

How better to police research

A draft NIH report rightly rejects suspicions of fraud at the Whitehead Institute, but by its deficiencies points to the need for the more effective self-regulation of research.

THE long and corrosive 'Baltimore affair' is nearing its end, but it is still not clear what the end will be. The affair is that engendered by the publication in April 1986 of a complicated article on the immunology of transgenic mice in the journal *Cell*, and is misnamed. Dr David Baltimore, the director of the Whitehead Institute at the Massachusetts Institute of Technology (MIT), happens to be the best known of six authors and the one who has voluntarily shouldered the burden of dealing with the controversy surrounding that single article. The controversy itself springs from assertions that the data reported in the paper do not sustain the conclusions drawn from them, in itself an unremarkable circumstance. But the assertions, and the denials they have provoked, have been invested with enormous significance. Investigations have been mounted by two universities (MIT and Tufts), the National Institutes of Health (NIH) and, most recently, the US Congress. At times, it has seemed that the correctness of this single scientific paper would represent, in the public mind, the integrity of the whole scientific process. More immediately, there is a risk that the controversy will help provoke legislation in the US Congress that imprudently restrains research.

The technicalities of the argument are intricate (see *Nature* 333, 795; 30 June 1988) but boil down to the question whether the immune system of a transgenic mouse is unexpectedly prone to produce antibodies sharing the specificity of those encoded by the transplanted gene. The human side of the story is simpler. Dr Margot O'Toole, a postdoctoral fellow at MIT's Cancer Research Center who was temporarily assisting Dr Thereza Imanishi-Kari, one of the principal authors of the disputed article, objected to the quality of some of the data gathered in the laboratory. Her charges were investigated by MIT and eventually dismissed. O'Toole has since left research.

Complaints

Subsequently, O'Toole's complaints were taken up by the twin-like combination of Dr Ned Feder and Mr Walter Stewart, employees of NIH acting in this connection as independent scientists, and considerably elaborated. Neither O'Toole nor Feder and Stewart have accused the authors of fraud. The essence of O'Toole's complaint is that a reagent used for characterizing immunoglobulin molecules is less specific than assumed in planning the experiments and in interpreting the data gathered from them. Feder and Stewart have included in a widely circulated unpublished manuscript an explicit statement that they do not allege fraud, but their criticisms have often been rejected as those of persons of no standing, either as immunologists or collaborators. In a widely circulated letter to colleagues in May this year, for example, Baltimore explained his and his associates' decision not to supply Feder and Stewart with further information by saying that "they are not immunologists" and that they are "self-appointed and had no right to the data". But eventually an expert committee was appointed by NIH; there is a summary of its draft report on page 505. Meanwhile, the issue has been taken up by the oversight subcommittee of the US House of Representatives' Committee on Commerce and Energy, which has held hearings on the issue and may pursue it

further after 20 January. The status of the draft NIH report should not be misunderstood.

The document has been sent for comment to those principally involved, may be modified in the light of what they say and will eventually be published together with those comments and whatever gloss NIH chooses to put on the affair. The draft insists that there is no evidence of "fraud, misconduct, manipulation of data or serious conceptual errors" (which is exactly what reasonable people have known all along). But the panel also draws attention to a number of invalid assumptions and procedural errors in the conduct of the research. So much has been partly acknowledged by the authors, four of whom contributed a partial retraction to *Cell* two weeks ago. But the NIH panel's draft also says that the retraction is insufficient. The weeks ahead will no doubt see further argument on this point, but the exact balance of right and wrong is now irrelevant to the more general issues raised in the past two years.

Exoneration?

The obvious question is that of how all parties to the dispute can apparently be exonerated by the draft report. O'Toole's chief point, that at least one of the serological reagents used in the experiments is not nearly as specific as had been claimed, was acknowledged by Baltimore in May and is covered by the retraction. Some of Feder and Stewart's more varied criticisms are also seen to have a force not previously acknowledged. But the authors are said to have done an honest and comprehensive job.

The panel is admittedly more cautious about the validity of the main conclusion than was the authors' retraction, which says flatly that it is "unaffected"; since the original publication, there has been published evidence both for and against this conclusion. Those working in the field and wishing to extend what has been done already will probably be helped less by the general propositions so far advanced than by the warnings provided by the original *Cell* paper of experimental difficulties ahead.

The apparent contradictions of these conclusions are not as remarkable as they seem — and will undoubtedly be represented to be. The truth is that the original experiments, while interesting as one of the first sets of observations of the behaviour of immunoglobulin genes in transgenic mice, have also been shown to be not particularly decisive. Knowing what they now know, the authors would probably have designed them differently.

Even so, and despite the errors catalogued by the NIH panel, it does seem that endogenous immunoglobulin genes in transgenic mice are influenced by the transplanted gene even when it is not active (expressed). To what extent, and why, remains to be discovered. Even the original paper advanced two explanations, since when there have been others. What the world, and the US Congress in particular, must understand is that much of science is like that. The rare memorable discoveries are those that make scales fall from people's eyes; most others log observations that seem to make little sense. The decade of accurate atomic spectroscopy culminating in 1926 with the quantum theory was full of discoveries of the second kind.

So who are in the right? Baltimore and his fellow-authors? Or their critics? The NIH panel does not directly deal with this



question, perhaps because it, too, cannot appreciate how the answer can be "All". But that is what its report means.

Moreover, it is important that the credibility of this surprising answer should be widely appreciated, for similar questions will arise again. Baltimore and his supporters (which means the bulk of the scientific community) would no doubt prefer that the inference from the NIH panel's report should be less equivocal.

In the year of the "sound-bite" (the half-sentence slogan of a candidate for the US presidency), "right" is an appealing concept. But it is inappropriate. There is no fraud, and there never was one; the general character of the authors' original conclusions may be correct; but the experiments and their interpretation are flawed by errors of various kinds, some of them downright sloppiness, as the critics have been saying.

How will this affect the public reputation of research, not to mention the chances of congressional regulation? Shocking though it may seem, the content and even the conclusions of the original paper are probably irrelevant. But it may be important that the research community, which knows well that there is no such thing as absolute truth, but only successive approximations thereto, should now make propaganda for the general belief that not all discoveries shake the world, but may be none the less valuable for that. Yet it will be a matter of public importance that even honest work at a renowned laboratory in the United States should be discovered to be so error-prone. The more substantial errors stem from the use of biological assays whose validity had not been fully established. Others, of a clerical nature, seem trivial. But what will be made of the two measurements shown by the laboratory notebooks not to have been made but reported as "one . . . positive and the other negative"?

Accountability

The seemliness with which this dispute has been conducted will matter more to the outside world, congressional committees included. The record is not particularly encouraging. O'Toole's objections were overridden, but have been at least partly validated. Feder and Stewart, whatever their status, were dealt with haughtily at the outset and have since had a rotten time. They deserve an acknowledgement that their questions have not been mischievous. True, with the passage of time, the original authors have also suffered. It is especially shabby that Baltimore in particular, one of the most distinguished of his generation in the United States, has been pilloried in the general press. Personality conflict, misunderstanding and sheer pride may be part of the trouble, but the research community has a nasty situation. There is the issue of accountability. The simple theory that researchers who spend public funds should be accountable to the public for every detail of what they do is, of course, a nonsense. It would not work. Not even the US Congress would be a competent judge of what is proper, while researchers would follow other careers if it tried to be. The legitimate public interest is merely that public funds should be spent responsibly, which means that grant-making agencies should be able to give coherent accounts of their policies and that individual researchers should be able to give an equally good account of themselves when asked.

While the constitutionally proper invigilators are the grant-making agencies themselves, circumstances in the United States are peculiar in that the Congress is obliged to listen to all whistle-blowers, the number of whom has been magnified by the attested cases of malpractice accumulating in the past ten years. Researchers can no longer disdain enquiries that would have seemed impertinences just a few years ago. But researchers who are challenged should not themselves have to shoulder all the burden of their own defence. The past two years at the Whitehead Institute will have demonstrated all too clearly how disruptive is this process, and how unsatisfactory.

Experience (of which there has recently been too much) has also shown that internal inquiries are often favourably biased while inquiries mounted by public grant-making agencies (such

as NIH) too often produce reports that, by a quasi-judicial balance of contradictory factual statements, fail to answer the questions the whistle-blowers ask. This week's draft report is a good example. That is why the research community needs a new system for creating friends at court. The ideal would be that the National Academy of Sciences (which has had too little to say about the Baltimore affair) should lend its prestige and influence to a more seemly solution of recurring problems. The ideal would be a large panel of the academy's members from whom could be selected small groups able to inquire quickly and convincingly into fresh charges of impropriety as they arise, willing to stand by their conclusions (favourable or otherwise) and ready to champion the belief that honest mistakes are unavoidable and frequent (which is the truth).

Finally, there is a matter of professional manners to be dealt with. At difficult times, people tend to lapse into reticence when articulateness would be their best self-defence. To urge that whistle-blowers should be given a more sympathetic hearing than they usually get is not to give every nit-picking graduate student a licence to subject anybody in sight to an inquisition, but is merely to suggest that successful courteous intellectual argument is, in the long run, the best way of disarming invalid allegations of misconduct. It is naturally essential to avoid the suspicion that whistle-blowers on that account suffer serious career hazards (but, naturally, people disappointed in research are more likely to look for flaws in what others do). But these would all be ingredients in a serious attempt by the scientific community to regulate its own affairs which, in the United States, is the best means of making sure that the Congress does not attempt the job instead, causing mayhem in the process. Baltimore and his fellow-authors might usefully make a start by responding frankly, and more generously than in their *Cell* retraction, to the draft NIH report. □

Copy rules to copy

An agreement between IBM and Fujitsu may bring order to a disordered market.

THERE is much to be learned from the arbitration settlement last week of the long-standing dispute between International Business Machines (IBM) and the Japanese computer manufacturer Fujitsu. Most obviously, the size of the settlement (\$833 million over 15 years) is a vivid measure of the value that may be attached to intellectual property. Although there are some who hold that Fujitsu might justly have paid more, \$800 million is serious money; its amount should give pause to manufacturers seeking to copy others' software.

More constructively, the settlement should help define rules for telling what is permissible in the software business. The circumstances are that Fujitsu had copied the software constituting the operating system of a range of IBM's mainframe computers and was then challenged with impropriety. The status of protection for software was, and still is, muddy. The patents procedure does not apply. Software can (and is) protected by copyright legislation, but precisely how is often far from clear.

The importance of last week's arbitration settlement is that it defines where the line should be drawn in at least one important case. Fujitsu has been making annual payments to IBM since 1983 in acknowledgement that it owes something; in return (and properly) it has looked for a continuing right to amend its copied program as IBM's products are themselves amended and improved. The nub of last week's agreement determines how that should be done. IBM undertakes to let Fujitsu know, and even to discover for itself by inspection, the functions its programs have. But Fujitsu must then develop software of its own to perform these functions. The corresponding duty laid on IBM is that it should not deliberately obscure the functions of its software so as to make it impenetrable which, given that money is changing hands, is not unfair. □

No misconduct or fraud in Baltimore case

- NIH panel circulates draft report
- Errors and lapses in judgement recorded

Washington

THE panel of enquiry set up to investigate possible "misrepresentation or misconduct" in a paper published in the journal *Cell* by Thereza Imanishi-Kari, Nobel laureate David Baltimore, and four others, records "significant errors of misstatement and omission, as well as lapses in scientific judgement and interlaboratory communication" in its draft report to the National Institutes of Health (NIH). But the report emphasizes that "no evidence of fraud, misconduct, manipulation of data, or serious conceptual errors" was found.

The draft report, written by Joseph Davie of Searle Pharmaceuticals, Hugh McDevitt of Stanford University and Ursula Storb of the University of Chicago, has not been publicly released but has been circulated for comment amongst all those involved in the case. An NIH spokesperson says it hoped to make its findings public before the end of December.

The panel of enquiry was asked to investigate only whether the *Cell* paper is "scientifically accurate", whether there is evidence of "misrepresentation" and whether there is a need to publish "appropriate corrections in the scientific literature". It did not deal with the controversial issue of whether complaints made about the paper two and half years ago by Margot O'Toole, a postdoctoral student in the principal author's laboratory, were dealt with fairly. O'Toole claimed that she had been portrayed as making "ridiculous, trivial and unsubstantiated complaints against my superior for vindictive reasons" and had felt forced to abandon her scientific career.

O'Toole's complaints attracted public attention after her case was taken up by Walter Stewart and Ned Feder, two NIH scientists well known for their controversial studies of scientific misconduct. Their attempts to publish a study analysing the *Cell* paper helped spur investigations both by NIH and the congressional energy and commerce subcommittee on oversight and investigations, chaired by John Dingell.

The draft reports' chief criticism is that inaccuracies in Table 2 of the paper are "sufficiently serious to merit a letter to *Cell* informing the editors of this fact". A corrective note recently published in *Cell* by Baltimore and three others is said to "not fully meet with the recommendations of the panel". The publication of the corrective note had already been attacked

by Dingell as an attempt to pre-empt the findings of the NIH enquiry.

The conclusions of the NIH panel differ from those reached by two earlier informal investigations, one organized by Henry Wortis of Tufts Medical School in Boston and the other by Herman Eisen of MIT. Although some fault was found with the paper, neither committee found a published correction warranted.

The NIH panel accepts several criticisms of the paper. The paper had sought to examine the effect of a rearranged exogenous immunoglobulin gene on the expression of the host mouse's immunological repertoire. While the panel agreed that a key reagent, Bet-1, could discriminate between the products of endogenous gene and the transgene, they found that the sensitivity of the Bet-1 assay was not sufficient to rule out alternatives to the paper's main conclusion. The panel agreed that the value of a cut-off point, above which binding was considered positive, was arbitrary and "might not have been appropriate". The panel found most fault with Table 2 where it noted that "isotyping was not done, but was claimed to have been done" and that the data are for "wells, not clones".

With some of the scientific controversy surrounding the paper apparently ready for public airing, attention may turn to the wider issues surrounding the case that were raised by the Dingell committee. Dingell's aim was very different from that of the panel appointed by NIH. He set out not to examine the "the scientific merits of the case" but to examine "how the system handles concerns or allegations of possible scientific error or misconduct." His concern was spurred by a rash of misconduct cases in which he claimed that "NIH has unwittingly aided in the harassment of whistle-blowers" and by concern that in the two years since the Baltimore case began to unfold, NIH had "done nothing".

It is not yet clear how many of these broader issues will be dealt with in the final version of the NIH report, or even whether the NIH committee will deal with the scientific issues to everyone's satisfaction. O'Toole this week confirmed that NIH had lived up to their assurances and allowed her to comment on the draft report. But unless there are dramatic developments soon, it seems likely that the congressional investigation into the circumstances surrounding the case will continue early next year.

Alun Anderson

Pressure on UK for embryo bill

London

As the British government continues to delay the introduction of legislation on assisted fertilization and embryology, both those bodies which support research and those which oppose it are becoming frustrated. And the workload and financial pressures on Voluntary Licensing Authority (VLA), currently responsible for monitoring programmes of in vitro fertilization (IVF) and embryo research, are increasing. The authority is currently funded by the Royal College of Obstetricians and Gynaecologists and the Medical Research Council. But the RCOG is now to retract a substantial proportion of its funding and the MRC cannot afford to increase funding says Dame Mary Donaldson, chairman of the VLA. The authority is now about to request funding from the Department of Health to meet the shortfall; those funds would only be in support of the authority's clinical activities and not for research.

Donaldson says the authority is "pretty exasperated" by the government's delay in legislating on embryo research. When the authority was set up in 1985, it was told it would be in operation for only one year. Since then its lifetime has been extended annually, and it may be around for another three. Now the authority wants its name to be changed to the Interim Licensing Authority, partly as a reminder that a statutory authority is expected.

The restrictions on resources are preventing the authority from carrying out its duties as it would like. Ideally each license would be reviewed annually, but this is not possible at present. And the increasing use of the GIFT (gamete intra-fallopian transfer) technique is also stretching the authority's resources, says Donaldson.

Another body frustrated by the delay in action by the government since the white paper (policy document) was produced last year is the Roman Catholic Bishop's Conference of England and Wales. The Archbishop of Westminster, Basil Hume, has written on behalf of the conference to the prime minister expressing dismay that a bill was not announced in the Queen's Speech, when the government's main plans for the new parliamentary session are outlined. The Bishops plan to keep up pressure on the government to introduce a bill this year.

In the government's white paper, two alternative proposals were made, one banning research, the other permitting limited research on embryos up to 14 days old (See *Nature* 330, 409; 1987). The government said Parliament would be given a free vote to choose between those options.

Christine McGourty

West German science council out on strike

Munich

THE Wissenschaftsrat, founded in 1957 as an independent advisory council on science to the West German federal and *Länder* governments, has for the first time in its history gone on strike. The 22-member scientific review commission walked off the job at a meeting in Berlin in mid-November and is unlikely to resume its duties until next year.

The group walked out to protest at a budgetary manoeuvre of the West German Bundestag (parliament) that would move the position of general-secretary of the science council out of the (independent) West German president's office into the (partisan) federal education Ministry. The scientific members called the move "an attack on the independence of the Wissenschaftsrat as a whole".

The current troubles began when the Bundestag budget committee reduced the position of general-secretary from "B9" — a political appointee earning as much as DM11,000 per month — to "B7" — a public servant earning about DM1,000 less. But the office of the president refused to allow a B7 position in its budget, as such an employee has lifelong tenure. So the position ended up in the education ministry. The decisive blow came on 10 November, when an education ministry official informed the council that the position had been reduced even further, to B6, and that it would remain in the education ministry. The scientific commission decided unanimously an hour later to strike.

The commission, which is scheduled to meet again next Monday, the 12 December, has declared its willingness to compromise in case the general-secretary's position cannot be reassigned to the office of the president. The *Länder* education ministers have promised to support the striking members of the council in negotiating with their respective minister-presidents.

In the meantime, scientific review of several institutions has had to be postponed, among them the German Cancer Research Centre. Even more importantly, the necessary Wissenschaftsrat approval could not be given for fourteen new *Sonderforschungsbereiche* (special collaborative programmes) of the Deutsche Forschungsgemeinschaft, nor for a DM500 million university investment plan.

The earliest possible date for a settlement is 15 December, when the minister-presidents are scheduled to meet in Hamburg.

Steven Dickman

National Institutes of Health stop Harvard researchers' grants

Boston

DRAMATIC action marked a widely publicised case of alleged scientific misconduct last week, when the National Institutes of Health (NIH) terminated hundreds of thousands of dollars in grants to Scheffer C. G. Tseng and Kenneth R. Kenyon. The two scientists, both affiliated to Harvard University, are alleged to have profited from business interests in a drug that they were responsible for testing and to have delayed publication of data showing that the drug was ineffective.

But the severity of the NIH decision was soon under fire. John Dingell (Democrat, Michigan) was "horrified", according to a staff member on the energy and commerce subcommittee on oversight and investigations which Dingell chairs. Although the subcommittee, which is conducting its own investigation into the case, has previously criticized NIH for dragging out their investigations of scientific misconduct, the subcommittee thinks NIH have acted too hastily this time. As well as the congressional inquiry, two separate NIH investigations are still under way.

NIH cut off funding for the final two years of a five-year, \$850,000 grant to Tseng, now a research scientist at the University of Miami. According to an official Harvard statement, the conflict of interest arose in 1985 when Tseng, a former Harvard scientist, and his supervisor, Kenyon, acquired stock in a company that owned the rights to a Vitamin-A based ophthalmic ointment whose efficacy they were testing (see *Nature* 335, 754; 27 October 1988). NIH also suspended consideration of a

Harvard resignation

Boston

In a second case of alleged scientific misconduct at Harvard University, Dr Shervert Frazier, psychiatrist at Harvard Medical School and former director of the National Institute of Mental Health, resigned after admitting that he had plagiarized large sections of four review articles published between 1960 and 1975.

Frazier's decision to resign has upset colleagues, who consider resignation a stiff penalty given that none of the plagiarized material reported original research data.

Daniel C. Tosteson, dean of Harvard Medical School, emphasized to reporters that Harvard's rules on plagiarism "are pretty harsh". Tosteson says that Harvard guidelines to students state explicitly that plagiarism will mean expulsion. "Are we to say that it's OK for a professor?" he asks.

Seth Shulman

\$191,000 grant application submitted by Kenyon and insisted that he be replaced as the director of a NIH-funded postdoctoral fellowship programme at the Eye Research Institute, a private organization associated with the hospital. Kenyon has held the directorship for almost a decade.

Jack McLaughlin, associate director of the National Eye Institute of NIH, stresses that the cessation of funding represented "no presumption of any wrongdoing" by Tseng and Kenyon. But Tseng's lawyer, Robert Kasky, argues that NIH's termination of funds is "unreasonable, unjust and unfair" given that "the investigation of alleged improprieties is not nearly over". Kasky is not alone in his concern. Charles L. Schepens, Kenyon's colleague and president of the NIH-funded Eye Research Institute in Boston, likened the suspension of funds to the "mass hysteria" of the McCarthy era.

But Katherine L. Bick, NIH deputy director for extramural research, disagrees. "We felt it was proper", she says, "to simply stop where we are until we know what is going on." She emphasizes that the move does not reflect a policy change and is not without precedent. On other occasions, she says NIH have decided to "not make new awards" when questions of impropriety have been raised.

Before NIH took action, Daniel C. Tosteson, dean of Harvard Medical School, disclosed that Harvard's internal investigation found Tseng and Kenyon had never reported the potential conflict of interest. According to Tosteson, "proper research procedures were not followed" when the researchers altered their study's original protocol without notifying Harvard's review board. Tosteson states that he is "concerned about the damage this incident does to the medical community", adding that "such episodes threaten the public trust and the integrity of medical research".

Meanwhile, in a more recent development, an investor has sued Tseng, Kenyon and others, including Johns Hopkins researcher A. Edward Maumenee, and officials of the drug company in question — Spectra Pharmaceuticals Services, Inc. — for defrauding him in his purchase of Spectra stock. The class-action suit, filed in a US district court, alleges that the scientists presented a distorted and misleading picture of the eye ointment's effectiveness. Claiming that the entire future of the small company was predicated upon the drug, the suit seeks retribution on behalf of the hundreds of people who purchased stock in the company based on the allegedly misleading information.

Seth Shulman

FBI encourages librarians to be on the look out for spies

Washington

MUCH to the disappointment of libraries around the United States, the Federal Bureau of Investigation (FBI) will continue its controversial Library Awareness Program which seeks the assistance of library workers in identifying foreign spies using libraries.

The FBI programme became a national issue last year following an agent's visit to a Columbia University library. The FBI programme involves certain scientific and technical libraries (including university and public libraries) in the New York City area, and is an attempt to inform librarians of "intelligence-gathering activities that may be harmful to the United States", and to enlist librarians' help in identifying those activities.

The FBI focused on New York because of the large number of foreign missions located there. Library visits in other parts of the country, the FBI maintains, were connected with specific investigations.

But librarians have argued that the programme as initially conceived was far too vague. How, they asked, were they to determine which "intelligence gathering" was harmful? And how was it possible to identify library users as Soviet or Soviet-bloc nationals? Was a foreign-sounding name a sufficient criterion?

The FBI's request also raised the question of confidentiality of library records. Thirty-eight states now have library confidentiality laws, and libraries tend to look askance at steps that might curtail users' access to information in the library, a right protected by the first amendment of the US Constitution which

guarantees freedom of speech.

Representative Don Edwards (Democrat, California), chair of the civil and constitutional rights subcommittee, held hearings earlier this year on the FBI's activities. Following the hearing, FBI director William Sessions wrote to Edwards, outlining more specific guidelines on how the programme would function. Librarians will be told to look out for



individuals who identify themselves as Soviet or Soviet-bloc nationals and seek assistance in conducting library research, request referrals to students or faculty who might assist in research projects, remove library materials without permission, or "seek certain biographical or personality assessment information from librarians themselves and/or on individuals who are known to the librarian being queried, particularly on students and academicians".

University problems with database users

Berkeley

UNIVERSITIES may have to worry about keeping an eye on their database users. Some databases come with a restriction that bars access to non-US citizens. But librarians remain unclear over the extent of their legal liabilities.

The problem attracted notice last spring when Gary Lawrence, University of California coordinator of library affairs, directed librarians throughout the university to stop using the National Aeronautics and Space Administration (NASA)'s RECON aerospace database after discovering that the use agreement barred foreign citizens from access to it. He advised libraries to switch to the Dialog database which includes much of the same material as RECON, and costs more, but has no citizenship requirement.

But John Wilson, manager of database products for NASA, says the US citizenship

clause in the RECON agreement is misleading. The true intent is that the information — which is subject to the Export Control Act for purposes of maintaining US technological superiority — should not be exported to foreign governments or companies. Restrictions on the use of citations and abstracts are the same, says Wilson, whether they are retrieved through RECON or through Dialog. He says that in neither case is US citizenship a requirement for access, as long as the information will be used in the United States, for research or educational purposes.

That explanation does not convince Lawrence. He maintains that by signing the agreement, with its citizenship requirement clause, the university would expose itself to legal liability should a non-US citizen violate the export restrictions.

Marcia Barinaga

Software exports

Bangalore

THE government of India is to launch a package of incentives designed to boost the export of computer software. Most conspicuous is the provision of duty concessions on computer hardware imported for the development of software. The Department of Electronics is to complete technology parks at Bhubaneswar, Bangalore, Chandigarh and Pune geared to the export of software.

Dedicated satellite channels will be provided for computer links with Western countries. Indian software exports, mainly to the United States, are currently worth Rs800 million and are expected to be worth as much as Rs3,000 million by the end of the decade.

The software trade is controversial because US engineers protest that they are losing jobs to India, and Indian engineers protest that they are much less well paid than their US rivals. Radhakrishna Rao

Sessions promised that the FBI would not ask for information about people with "foreign-sounding names or accents", or for reports of "suspicious" or "anomalous" behaviour. In his letter, Sessions also said that if librarians do not wish to help the FBI in this effort, "that is up to them".

James Schmidt, chair of the American Library Association's Intellectual Freedom Committee, in a letter earlier this month to Edwards, agrees that Sessions has narrowed the scope of the programme, but believes that the FBI still does not understand that "however important [the FBI's] duties may be, they are subordinate to the first amendment rights of patrons lawfully using a library, and to state confidentiality laws".

A congressional attempt to enact federal confidentiality laws protecting library users was made this year, but withered under a blistering attack from the FBI. The protections were to be added to the Video Rental Privacy Protection Act (now law) that protects video store rental records. The act resulted from outrage over an article in a Washington tabloid describing the movies that Supreme Court nominee Robert Bork had rented. The article appeared while the Senate was considering the Bork nomination.

When Congress attempted to protect library records as well — using the logic that it was perverse to protect those who wanted to watch *War and Peace*, but not those who wanted to read it — the FBI objected. It insisted that a clause be added giving it access to library records under the aegis of a national security letter, an instrument that requires no judicial supervision. Threatened with this, library groups withdrew support for the bill, opting for the status quo. Joseph Palca

Coordination is the key for new AIDS research programme

Paris

ON the eve of World AIDS Day, the French research and technology minister, Hubert Curien, unveiled details of a new pluridisciplinary programme to back AIDS research. Criticized earlier in the year for a lack of positive action and for underfunding (see *Nature* 333, 103; 1988) Curien says money is no longer an issue. The new programme has a budget of FF 150 million (£14.19 million) but, he says, "if researchers say they need more money, I have the assurance of the minister for finance and of the Prime Minister that extra funds will be available" — a promise that won a smile from Luc Montagnier, author of the earlier attack.

The emphasis of the programme is on national and international coordination. All AIDS research now comes under the umbrella of a new AIDS agency, headed by professor Jean-Paul Lévy and advised by a scientific advisory council and a series of specialist panels.

The budgets available for basic research represent a doubling of last year's levels. The programme also includes, for the first time in France, a budget (FF3 million) for sociology research on AIDS and FF10 million for research into prevention.

The inclusion of a human sciences budget reflects a new awareness in political circles that there is no hope of a cure or a vaccine in the immediate future. Meanwhile, little is known of the behaviour of populations at risk nor of the projected socio-economic effects of AIDS. As change of behaviour is the only effective prophylactic likely to be available for some time, says Curien, it is essential that information campaigns are better targeted. One of the chief responsibilities of the new agency will be to encourage

researchers "on the fringes of AIDS research" — sociologists, sexologists, and epidemiologists.

To ensure that the results of research are shared between laboratories and that they percolate rapidly to the public, the agency will organize two annual national symposia, regular meetings on specific topics and the publication of a three-monthly information bulletin. Regular exchanges and joint projects with foreign laboratories will also be reinforced. Since last year, a formal programme of cooperation has existed with West German laboratories, while the UK Medical Research Council has been collaborating with French counterparts. An Anglo-French clinical trial of AZT has just started. While the will to stimulate research is clear in government's programme, it will nevertheless be constrained by a shortage of qualified researchers in some areas.

Peter Coles

Condoms condemned

Paris

AN influential French consumer magazine has told readers that only 6 of 41 brands of condom tested would provide adequate protection. Following the publication in *50 Million de Consommateurs* of tests out by the national consumer institute (INC), the French Secretary of State for consumer affairs, Veronique Neiertz, ordered five brands of condom to be withdrawn.

According to Health Department statistics, only 9 per cent of couples in France use condoms, compared with 50 per cent in Canada and 68 per cent in Japan. Yet the Health Minister, Claude Evin, sees wider acceptance of the condom as the most effective means available to reduce sexual transmission of the AIDS virus.

The INC report subjected condoms to two tests of elasticity — inflation and elongation — and a test of porosity. Fourteen of the brands were judged to have failed the porosity test. The aim of the INC article, according to its authors, was to enable the French public to "choose the right."

Peter Coles

AIDS predictions until 1992

London

AT least 10,000 new cases of AIDS are bound to be diagnosed in England and Wales between 1988 and 1992, according to the predictions of an expert group published* last week on the eve of World AIDS Day, 1 December. While considerable uncertainties prevent any very definite predictions, the group is certain that the past exponential growth in cases will not be continued over the next few years. But, it concludes, "it would be a gross error to regard even the lower predictions as grounds for complacency".

Continued exponential growth would predict around 10,000 new cases in 1992 alone. Instead, the group estimates there will be only 25–30 per cent of that number. For the basis of planning, a figure of 3,600 is recommended. In fact, there are likely to be some 20 per cent more cases that are not reported; the group recommends an urgent investigation of the extent of under-reporting, which could be as great as 40 per cent.

The group, chaired by Oxford statistician Sir David Cox, was asked by the Ministry of Health to predict the numbers of HIV-infected individuals, and AIDS cases and deaths over the next 2–5 years as an aid to health-care planning. One major problem it faced is the lack of reliable figures on the number of people currently infected with HIV, from whom almost all the new cases of AIDS over the next five years will be drawn. The attention drawn to this problem in the report was one of

the factors that influenced the recent government decision to instigate involuntary and anonymous screening of blood samples for HIV antibodies as soon as the Medical Research Council has drawn up plans (see page 413 of last week's issue).

In the absence of direct information, three different methods of estimating prevalence were used and all suggest that between 20,000 and 50,000 people were infected with HIV by the end of 1987, of whom 6,000–17,000 are heterosexuals. Although this provides the potential for considerable further heterosexual spread, Cox and his colleagues assume, for their predictions, that there will not be a large increase in the number of heterosexual cases of AIDS in the next few years. They also assume that the changes in behaviour among homosexuals that have stopped the exponential growth in AIDS cases will continue into the future.

The report recommends that its predictions should be updated annually by the Communicable Disease Surveillance Centre, and the Ministry of Health has agreed — without immediately offering any extra resources. But the ministry has delayed any specific response to most other recommendations. These include an urgent need for the study of sexual behaviour in high-risk groups as well as investigation of factors affecting HIV transmission between intravenous drug users. Several suggestions for future epidemiological studies of infection may re-emerge as part of the Medical Research Council's plan of action.

Peter Newmark

World AIDS Day

Washington

LAST Thursday, 1 December, was World AIDS Day, sponsored by the World Health Organization (WHO) to focus attention on the global effort to stop AIDS. More than 100 countries put on special activities intended to increase awareness of AIDS and to discuss how to combat the epidemic.

World AIDS Day saw the number of reported AIDS cases rise to 129,385, a little over 60 per cent of them in the United States. But WHO officials fear that the real number may be twice as high, given the lack of accurate statistics from Africa. Information campaigns were launched in 39 African countries, where those most at risk of AIDS in Africa are young urban workers who make the biggest contribution to their countries' economies.

Lisa Perlman

* Short-term prediction of HIV infection and AIDS in England and Wales (HMSO, £6.50).

Sidney Drell resigns from Stanford University centre

Berkeley

SIDNEY Drell, the distinguished physicist and arms-control expert, has resigned as co-director of Stanford University's Center for International Security and Arms Control (CISAC), out of frustration over the university's refusal to grant faculty positions at the centre. Drell's leaving will weaken the technical dimension of CISAC's programme, may jeopardize funding and spur an exodus from the centre.

CISAC grew out of an arms-control programme begun in 1970 by Drell, physicist Wolfgang Panofsky and others, and has become a world-famous nuclear arms

policy centre under the co-directorship of Drell and Stanford political scientist John Lewis. Most of the 70 CISAC members are Stanford professors, and are only peripherally involved with the centre. But a core of about a dozen senior research associates, including experts on treaty negotiation and weapons systems, work full-time. Stanford's policy of restricting faculty appointments to academic departments has thwarted efforts to secure faculty positions for these people. The result, says Drell, is an unjust situation in which CISAC members are denied professorships, while the fellows they train go on to faculty positions in competing programmes at MIT, Carnegie Mellon and Harvard.

Drell said he resigned because his discussions with Stanford had stalled. He will continue as deputy director of the Stanford Linear Accelerator Center and will pursue his arms-control work through private consultancies.

Stanford provost James Rosse regrets Drell's decision, but defends the university's policy of reserving appointment

power for departments that offer coordinated teaching and doctoral programmes. He says the authority to make faculty appointments could maintain some interdisciplinary programmes beyond their useful lifetimes.

Drell counters that Stanford's inflexibility is a threat to academic quality rather than a safeguard. Academic excellence, he said, is best ensured by conducting national searches, and by broad review of appointments, not by insisting on pigeonholing appointments within departments.

Drell's resignation has prompted the Carnegie Corporation, which has provided CISAC with \$1.8 million over the past five years, to withhold approval of next year's \$460,000 grant payment until Stanford provides an assurance that the centre can continue its scientific work in Drell's absence.

Drell warns that several of his younger colleagues, including former astronaut Sally Ride and nuclear engineer Ted Postol, are also likely to leave. Postol says those who leave will not be easily replaced, as there are few scholars with appropriate qualifications. The loss could drive Stanford "out of the business", he said. "It will go from the best programme in the country to no programme."

Marcia Barinaga

Atlantis succeeds

Washington

THE space shuttle Atlantis, carrying a secret military payload, was successfully launched at 9.30 a.m. last Friday, 2 December. The launch had been postponed from Thursday because of strong high-altitude winds, and similar conditions the next day relented only a few minutes before the end of the allotted launch period.

To hinder Soviet attempts to track the shuttle's flight-path, the exact time of blast-off had not been revealed beforehand, and shortly after launch Atlantis turned sharply northwards rather than flying east across the Atlantic Ocean.

The satellite carried by the shuttle was widely reported to be a high-resolution radar imager called Lacrosse, able to map the Earth's surface regardless of weather. In the 2½-year hiatus before Discovery's flight two months ago, the US Air Force had sent several reconnaissance satellites into orbit using Titan rockets, but in the shuttle's absence the ability of the Department of Defense to follow Soviet military activities had been curtailed.

Atlantis' return to Earth was to be announced with 24 hours' notice, but on Monday 7 December the mission was still in progress with no landing time officially released.

David Lindley

Broder for NCI

Washington

SAMUEL Broder, associate director of the clinical oncology programme at the National Cancer Institute, is expected to be named by President Ronald Reagan as the new head of the National Cancer Institute as soon as a necessary background check has been completed. Broder, best known for his work on the AIDS drug AZT, will succeed Vincent DeVita, who left in August. The institute is the United States' largest cancer research facility and has a budget close to \$1,500 million.

Alun Anderson

Engineering lab privatization postponed

London

THE privatization programme of the British government suffered a severe setback last week with the postponement of the sale of the National Engineering Laboratory (NEL). The sale is being delayed for two years while a team of management consultants undertake a study of the laboratory and supervise its restructuring.

The sale, which was announced in June, was planned because most of the work carried out at the laboratory was termed industrially relevant and the government is adamant that the cost of such work is borne by the private sector. Since the government called for commercial bids to be made within six weeks, the sale has been fraught with difficulties.

Nine organizations made bids for the NEL, which is estimated to be worth £30 million, and negotiations with a Glasgow-based research organization, YARD Limited, took place between August and October, but ended without agreement. The Institution of Professional Civil Servants (IPCS), the union which represents the 300 scientists and technicians at the NEL, says that those negotiations were ended after the Department of Trade and Industry (DTI) realized that it would be illegal to go ahead with the sale because the staff had not been consulted on the reasons for it. Another problem, says the IPCS, was that YARD would have been obliged to take on the current collective

bargaining agreements, but that would have been impossible because YARD is a non-union company. And another sticking point was reported to be the amount of funding to be provided by the government during the period of transition into the private sector. At present more than half of the NEL's annual £21 million funding comes from government: the remainder comes from private industry. Now consultants Touche Ross will carry out a five-month study of the NEL, which will be followed by an 18-month period when an advisory body supervises the restructuring.

The IPCS accuses the government of "wreaking mayhem" within the laboratory as a result of its hasty privatization plans, and says that already staff are leaving because of the uncertainty surrounding the NEL's future. Though the government says privatization will still go ahead, another option is to give the laboratory 'agency' status.

Depending on the arrangements surrounding the transfer of public sector institutions to agencies, this can mean anything from an institution remaining almost identical to a private company or, at the other extreme, to part of the civil service. Other candidates for agency status are the remaining DTI establishments: the Laboratory of the Government Chemist, the Warren Spring Laboratory and the National Physical Laboratory.

Christine McGourty

Hard times for DFG

Munich

THE Deutsche Forschungsgemeinschaft (DFG) announced last week that it will be able to fund only 14 of the 30 'priority programmes' (Schwerpunktprogramme) applied for by university researchers. Priority programmes support collaborative projects in social and natural sciences as well as engineering for up to five years.

Projects were approved by the DFG Senate on 28 November in fields ranging from market economics to "genome analysis and gene transfer in domestic animals". A further nine projects deemed worthy of funding were rejected because of a lack of money. A witness to the Senate session said it resembled an "oriental bazaar", with Senate members desperately swapping proposals, trying to support as many as they could. In the end they were satisfied that at least a few good proposals were funded.

Until a few years ago, all qualified proposals for priority programmes were funded. This is the first time that fewer than half of these priority programmes could be supported. The number of applications reached an all-time high this year. DFG officials once again warned of the discouraging effects on researchers if more government money does not become available.

Steven Dickman

Less school science

London

THE amount of time that pupils at UK secondary schools spend studying science is the subject of the latest controversy to follow from the 1988 Education Reform Act. A working group on science set up by Mr Kenneth Baker, Secretary of State for Education and Science, recommended that in the fourth and fifth years of secondary school, pupils should spend 20 per cent of their time on science and aim for two GCSE exam certificates. But Baker asked the National Curriculum Council (NCC) to draw up a plan allowing some pupils to devote only 12.5 per cent of their time to science and aim to get just one certificate. In its report published this week, the NCC has outlined such a plan, but recommends that pupils should aim for the fuller programme, as the reduced one is lacking in breadth. Baker is expected to make draft orders following the NCC's advice before next year.

Mr Jack Straw, education spokesman for the Labour Party, says that Baker's plan is an "easy, cheap, but short-term way out of mounting shortages of science and maths teachers". And he cites figures which show that most of the universities and polytechnics which commented on the proposals of the working group are opposed to Baker's plan.

Christine McGourty

Difficulties ahead for Japan's fifth-generation computer plan

Tokyo

JAPAN'S Institute of New Generation Computer Technology (ICOT) is about to make a final spurt towards its goal of building a fifth-generation computer. The institute unveiled the fruits of seven years' work at its international conference in Tokyo last week along with plans for a final three-year push. But critics wonder whether ICOT can reach its target or even if it is headed in the right direction.

The Ministry of International Trade and Industry (MITI) sent shock waves through the computing world in 1981 when it announced plans to develop an entirely new generation of "user friendly", intelligent computers. Although the computers are not yet in sight, ICOT director Kazuhiro Fuchi was brimming with confidence at last week's conference. He says the research has been "smoothly conducted so far" and ICOT is ready to "jump" into the final phase of the project.

MITI has requested ¥6,448 million (\$52 million) for ICOT in fiscal 1989, 12.5 per cent more than this year. And Fuchi hopes to get more than ¥20,000 million for the final three years of the project (1989 to 1991) which will concentrate on building a massive parallel inference machine (PIM) and accompanying software. PIM will contain 1,000 individual processing elements and be capable of 100 to 1,000 million logical inferences per second, far beyond the capabilities of present computer technology.

But many difficulties lie ahead. At the conference a system of 64 sequential inference machines linked in a network was on display but ICOT researchers admit that preparation of software for the system was much more complicated than anticipated. David Warren, an expert on parallel computing from the University of Bristol, says that the biggest technical barrier to be overcome is the development of software for PIM. ICOT's approach is for the programmer to specify explicitly how the computational problem and sub-problems should be divided up. But Warren points out that parallelism should be hidden from the programmer if the computer is to be "user friendly", otherwise it can be programmed only by experts.

One of the key features of the computer originally foreseen by ICOT was an intelligent man-machine interface that would allow communication by speech and drawings. But, according to Hideo Aiso, the conference chairman, image-processing and speech-recognition were dropped in 1985 because of budget restrictions and lack of manpower.

Another line of research that has disappeared involves the knowledge-based

system at the core of the machine. ICOT took the unusual approach of trying to develop hardware rather than software for the storage and retrieval of knowledge. And at ICOT's last international conference in 1984 a prototype relational database 'machine' called Delta was displayed with much fanfare.

Nothing has been heard of Delta since. A public relations officer for ICOT says a new Delta machine will be built in the final phase of the project. Warren says ICOT has clearly decided to pursue development of knowledge-base software rather than hardware. But virtually no work has been done on parallelization of knowledge processing so this will be a major challenge for the final phase.

Fuchi admits that there are many problems to overcome. But he says that by taking up the challenge "human beings will be forced to come up with an answer", though he says the chances of successfully developing PIM are just 50 per cent.

Many participants at the conference question whether ICOT has chosen the right approach by sticking rigidly to logic programming and parallel processing. Competing projects in Europe and the United States, such as ESPRIT (European Strategic Programme for Research and Development in Information Technology) and DARPA (Defense Advanced Research Project Agency), have adopted a much more diversified approach and are, for example, branching out into neural network computing.

Warren says, however, that ICOT's project should not just be judged on whether it achieves its goals. The project has succeeded in "plugging Japanese computer researchers into the world research community" and is making a substantial contribution to basic research. PIM may not be the best approach, he says, but given that ICOT must come up with a machine in three years, it is probably the best way to go.

■ Britain's Information Engineering Directorate (IED) and Japan's Institute for New Generation Computer Technology (ICOT) have reached an agreement that will allow IED researchers to carry out long-term research at ICOT.

The agreement, announced at the conference, will enable two IED researchers to spend from six months to one year in Japan. An earlier attempt to arrange cooperation between scientists involved in the UK's Alvey project and ICOT fizzled out because of problems associated with patent rights. Under the present agreement, intellectual property rights will be worked out on a "case-by-case" basis.

David Swinbanks

Wild chimpanzees upgraded to endangered species

Washington

As a compromise between the interests of conservationists and biomedical researchers, the US Fish and Wildlife Service announced last week that it will upgrade the status of wild chimpanzees from threatened to endangered. The change will increase protection for chimpanzees by requiring more documentation on their movement. But because the change applies only to wild chimpanzees, and not those in captivity, conservationists say it leaves a huge loophole for the sale of wild-caught chimpanzees to researchers.

Conservationists wanted all chimpanzees classified as endangered to help stop the depletion of wild chimpanzee populations and to control the number which end up in research laboratories. The Humane Society of the United States (HSUS) says the number living in the wild has dwindled to between 150,000 and 200,000.

The HSUS, the World Wildlife Fund, and primatologist Jane Goodall filed a petition with the Fish and Wildlife Service last year requesting the reclassification of the chimpanzee. They identified the international biomedical establishment as "one of the greatest threats to the continued existence of wild chimpanzees", because the demand for the animals for research encourages poachers to capture and sell wild animals. The HSUS estimates that between five and ten chimpanzees die during the capture of every infant.

The United States has been prohibited since 1977 from importing wild-caught chimpanzees under the Convention on International Trade in Endangered Species. But many African countries have not joined the convention or lack the administrative procedures to enforce it, and conservationists claim that wild chimpanzees continue to enter the United States, which contains 80 per cent of the world's captive chimpanzees.

The National Institutes of Health (NIH) oppose reclassification of the chimpanzee, urging that a federally funded study of wild chimpanzees in Africa be completed before a decision was made. Conservationists agree that a systematic field survey is necessary, but have alleged that NIH's request for a study is a delaying tactic, because it could take up to four years to complete. According to George Galasso, chairman of the US Public Health Service AIDS Animal Model Committee, the urgency for conducting an NIH-sponsored study is now gone, but NIH's Department of Research Resources will have to make the final decision to drop the project. Representatives from that office refused to be interviewed.

The AIDS epidemic has heightened

concern over the supply of chimpanzees because they are the only animals besides humans which are susceptible to infection by HIV. Five breeding centres in the United States — set up in the late 1970s after trade in chimpanzees was banned — house 350 chimpanzees, but are producing less than 50 young chimpanzees for research per year. Out of the total of nearly 1,400 chimpanzees in US biomedical facilities, NIH officials say only about 800 are virus-free. Many were involved in the testing of the hepatitis B vaccine.

Galasso's committee reviews all federally funded research proposals involving chimpanzees to prevent the duplication of experiments, and to maximize their use by sequencing them through experiments, reserving infection with HIV as a final use.

Early proposals for the NIH survey of wild chimpanzees reportedly mentioned the possibility of using wild-caught animals in AIDS vaccine research, if US breeding efforts did not produce enough. NIH has repeatedly denied considering using wild chimpanzees in AIDS research, and Galasso says the number of animals available for AIDS research is adequate. To ensure that NIH keep their word, Congress in the 1989 NIH appropriations bill forbids NIH to buy wild-caught chimps, or to contract with laboratories that do. **Carol Ezzell**

Forest defenders awarded 'Nobel'

Washington

SOME of those in developing countries who are fighting to preserve tropical rainforests have been honoured in the Swedish parliament. Sahabat Alam (Friends of the Earth) of Malaysia and Jose Lutzenberger of Brazil share this year's Right Livelihood Award, better known as the 'alternative Nobel prize'. Unlike their counterparts in the West, members of Friends of the Earth in Malaysia often risk arrest and imprisonment for anti-logging activities.

Especially mentioned in the prize citation is Harrison Ngau, a 28-year-old Kayan, who has led human blockades of logging roads in an attempt to halt destruction of tribal lands in Sarawak. Ngau is under virtual house arrest and is not expected to be allowed to leave Malaysia for the prize ceremony.

Lutzenberger, now 61, worked for a chemical company until 1972 when he resigned to lead a crusade against Brazil's agricultural policies. His efforts helped to increase the World Bank's environmental sensitivities and fathered the Brazilian environmental movement. **Lisa Perlman**

More funds for forest research

London

THE international effort to reduce tropical deforestation and the associated environmental degradation was speeded up last week when external donors of aid to developing countries agreed to double spending on forestry research by 1995.

Bilateral and multilateral donors, development banks, non-governmental organizations and researchers met near London to consider the report of an international task force on tropical forestry research, set up in February. The meeting endorsed most of the report's recommendations, including a call for an increase in funds for forestry research from the present level of about \$50 million to \$100 million annually. The report says that international expenditure in support of forestry development has roughly doubled over the past five years to about \$1,000 million, but research expenditure is still only 5 per cent of that. Drastic action is needed to raise the amount of research, says George Holmes, chairman of the task force. The problems which need to be tackled are enormous and the knowledge of how to deal with them is poor.

The report recommends expansion of research in five priority fields: trees in farming systems and watershed management; natural forest ecology and management; tree breeding and improvement; utilization and marketing; and policy and socio-economic issues.

This research could improve the productivity of trees and significantly increase supplies of fuel, fodder and other goods, says the report. This could benefit some of the 120 million people in the Sahelian and Himalayan regions who rely on trees for those products and some of the 3,000 million people who will face shortages of fuelwood by the turn of the century.

To plan and coordinate the research, the task force recommends the formation of an independent research council. But at the meeting last week, that recommendation was rejected. Instead, a forestry unit will be attached to the international coordinating body, the Consultative Group for International Agricultural Research, to help donors, national bodies and research establishments in evaluating priorities, mobilizing support, focusing research and exchanging information.

The absence of any international focal point at present compounds the problems caused by the state of national research strategies, which, says Holmes, are mostly appallingly bad. The task force report says that more political awareness of the issues and secure funding are needed.

Christine McGourty

Diagnosis of schizophrenia

SIR—The cautiously euphoric leading article¹ and the News and Views item² on papers on the localization or non-localization of a susceptibility locus for schizophrenia on chromosome 5 (refs 3,4) merit comment. Despite the extremes of the 'anti-psychiatry' tendencies of the 1960s and 1970s, whose major exponents, Laing and Szasz, you cite in your leading article, no materialist account of the condition(s) diagnosed as schizophrenia could doubt that it (they) must be associated with reasonably specific changes in brain biochemistry and physiology, and that, in particular environments, certain genotypes may well be more likely to express such changes. The problem, as many decades of neurochemical psychiatry has revealed, is to distinguish those changes which are necessarily, sufficiently and exclusively related to the schizophrenic diagnosis. This task is not helped by uncertainties in the diagnosis itself, and the epidemiological evidence which points to its sharply skewed class and ethnic distribution in Europe and the United States today. Hence the relevance of Kennedy *et al.*'s conclusion that at present it seems that there are different loci, possibly leading to different neurophysiological abnormalities, resulting in a final common pathway "phenotype for schizophrenia". One might better amend this to read in the plural, "phenotypes for schizophrenias" and having done so, to recognize that (a) there is not a one-for-one correspondence between particular genotypes and particular abnormal behaviours such as schizophrenia, and (b) rather than invoke the concept of phenocopy for a phenotypic condition which "mimics" an otherwise genetic condition (as Lander does) it would be better to refer to the genetic conditions which may sometimes be associated with particular phenotypic traits as genocopies.

It is this uncertainty that diminishes the strength of Sherrington *et al.*'s hope that "it may eventually be possible for psychiatrists to consider genetic counselling in families where chromosome 5 linkage can be reliably established". For genetic counselling to give meaningful advice to such families, it would be necessary to know in how many cases of chromosome 5 linkage schizophrenia does not occur. That is, while Kennedy *et al.* have shown that such linkage is not necessary for schizophrenia, Sherrington *et al.* have not shown that it is sufficient — a prerequisite for useful genetic counselling.

Moreover, your leading article and Lander's review accept too easily the view that the genetic studies enable one to reverse the traditional clinical method in which a disorder is diagnosed by its phenotypic (in this case behavioural) manifesta-

tions. Instead, you want the genetic measurement to determine whether or not a person is labelled schizophrenic or schizophrenic-sensitive (see Lander's view, contrary to that of Kennedy *et al.*, that it is genetic differences which make possible "splitting schizophrenia into a collection of distinct diseases"). This molecular reductionism, reminiscent of the ill-fated dexamethasone suppression test for depression, substitutes biochemistry for behaviour and genetic analysis for psychiatric judgement. It leads you both into a naively neo-darwinian view that, somewhere along the line, there might be adaptive advantages for genes which would otherwise be selectively eliminated, and to conclude that the genetic analysis will make it "feasible to seek means of designing drugs . . . which are more specific and more effective than those available". Your optimism may be justified, but historical precedent suggests that biochemical explanations tend to follow clinical and pharmacological interventions rather than vice versa.

STEVEN ROSE

Department of Biology,
The Open University,
Walton Hall,
Milton Keynes
MK7 6AA, UK

1. *Nature* **336**, 95 (1988).
2. Lander, E.S. *Nature* **336**, 105 (1988).
3. Sherrington, R. *et al.* *Nature* **336**, 164 (1988).
4. Kennedy, J.L. *et al.* *Nature* **336**, 167 (1988).

Nobel reaction

SIR—Your article on the Nobel prize for chemistry (*Nature* **335**, 752; 1988) wrongly suggests that the reaction centre can oxidize water and release oxygen, whereas it is anoxygenic. I should also say that it is incorrect to give the impression that the paper by G. Flemming *et al.* (*Nature* **333**, 190; 1988) was significant in identifying bacteriopheophytin as an electron acceptor. This fact has been known for a very long time based on the studies of many groups in the United States, particularly the work of Parson, Windsor, Holten, Schenck, Dutton and Fajer, as well as Shuvalov and colleagues in the Soviet Union.

JAMES BARBER

AFRC Photosynthesis
Research Group,
Department of Pure
and Applied Biology,
Imperial College,
London SW7 2BB, UK

Radon and smokers

SIR—There was a calculational error in the estimates of the risk of lung cancer due

to radon for smokers and nonsmokers in the recently published BEIR IV report¹, subject of a news item in your 14 January issue². The modifying term for age at risk was inadvertently omitted in the program used to prepare Table 2-4 in Chapter 2 and Tables VII-12 to VII-23 in Appendix VII. When this term is included, the estimated risks for lifetime exposure decline. For example, the corrected estimate of lifetime risk at 1 WLM/Yr is about 20 per cent smaller for smokers and about 25 per cent smaller for nonsmokers. The risk to exposed smokers relative to similarly exposed nonsmokers therefore becomes slightly larger, 10.5 rather than 10.1. The results for males and females in the general population (without regard to smoking status), presented in Chapter 2 of the report, are not affected by this error.

Corrected tables for smokers and nonsmokers of each sex can be obtained from the board at the address below.

WILLIAM H. ELLETT

Board on Radiation Effects Research,
National Research Council,
2101 Constitution Avenue,
Washington DC 20418, USA

1. *Health Risks of Radon and Other Internally Deposited Alpha Particle Emitters: BEIR IV* (National Academy Press, Washington 1988).
2. Palca, J. *Nature* **331**, 107 (1988).

Virulence opposition

SIR—All societies have their pariahs and the scientific society is no exception. We should promote research and science that is in mankind's best interests and condemn that which is contrary.

The letter entitled "Increased virulence of *Yersinia pseudotuberculosis* by two independent mutations" (*Nature* **334**, 522-525; 1988) shows that introduction of mutations in both the invasin gene and the Yop1 gene enhanced virulence of *Y. pseudotuberculosis vis-à-vis* the wild type; individual gene mutations diminished the virulence. *Y. pestis* causes plague. The authors introduced the functional Yop1 gene into *Y. pestis* and diminished its infectivity.

This work was carried out by members of the Swedish Defense Research Establishment. Sweden is a major supplier of arms to developing countries. It would be naive to think that the work was supported because it wanted to diminish the infectivity of the plague bacillus. When the Swedish Defence Establishment succeeds in introducing the correct form of the invasin gene as well as the Yop1 gene into *Y. pestis*, thereby producing a super-plague bacillus, do you think it will brag about it in the pages of *Nature*?

JACK A. KORNBLATT

Biology, Concordia University,
1455 de Maisonneuve,
Montreal, PQ,
Canada H3G 1M8

New ways with gravity waves

There is a chance that the United States will have a detector capable of recognizing gravity waves in the next few years. But the federal budget deficit may be a hindrance.

The California and Massachusetts institutes of technology (Caltech and MIT respectively) are traditionally proud of their differences, sometimes recognizable as rivalry, but they have for some months been breaking new ground by working together on a joint application to the US National Science Foundation (NSF) for what will be the next gravity-wave detector most likely to succeed. The timetable is leisureed, although that is not how it seems to those involved. They expect to have the financial gist of the application ready by next June, in time for the US budget cycle, and a more detailed specification of the equipment by September.

If President George Bush, the US Congress and the federal government's budget deficit allow, the budget for the fiscal year 1991 (beginning on 1 October 1990) may include an appropriation of \$150 million for a pair of detectors at opposite ends of the country (the Mohave desert and in Maine). There is talk of improving on that state of affairs with a third instrument elsewhere, Europe perhaps.

Each of the two institutions has had a vivid interest in the field since the early 1970s, when Dr Joseph Weber at the University of Maryland first demonstrated that gravity-wave detectors can be built and — less deliberately — how easily instrumental and spurious noise can be mistaken for genuine signals. Weber, it may be recalled, ran separate instruments on the campus of the University of Maryland, north of Washington DC, and at the Argonne National Laboratory, using a radio-link to compare the two signals (essential if local disturbances are to be eliminated). He explained at the time that he would have recorded the signals separately on magnetic tape if he had had the funds to pay for the stock of tape needed for a more objective comparison, much in the way that radioastronomy observatories collate measurements at separate telescopes. But the radio link could just as well have been the source of the coincidences that misled him.

It is interesting how solidly informed opinion has moved away from the instrument designs of a decade ago. Then, in the years after Weber's work, detectors were simply large symmetrical massive objects — aluminium cylinders, for example — suspended so as to pick up mechanical movements from the surroundings. Earthquakes but also passing traffic are a nuisance. The crucial measurement is that

of the displacement of the free surface of the suspended object, usually done by cementing a mirror to the surface and setting up an interferometer between that and an external source of light, naturally a laser.

The most obvious reason why detectors along these lines are out of fashion is mechanical. The sensitivity of a suspended object used for the detection of gravitational waves increases with its linear dimensions, but that complicates vibration-free suspension. Then there are unavoidable difficulties about the frequency spectrum of the detector, which is a complicated function of its geometry. In all the circumstances, it is remarkable that the sensitivity of the vibrating bars, which has steadily improved since Weber's first experiments (he is still active in the field), amounts to a relative strain of 10^{-18} . The enthusiasts look for further improvements, perhaps by a total factor of 100, but it seems to be agreed that there will then be a halt to the process, perhaps set by quantum noise (for which reason new designs of vibrating detectors are cooled to liquid helium temperature).

Luckily, it seems now to be agreed that there is a way forward to the sensitivity that may pick up gravitational waves from the distant objects in the Universe advertised to emit them with sensitivity in the range of 10^{-22} — colliding black holes and, more likely, exploding supernovae.

Logic (rather than NSF, which was unkindly reported by the *Los Angeles Times* last month to have decreed a "shot-gun marriage") seems to have forced Caltech and MIT in similar directions. Each team has settled on the principle of building a laser interferometer from light traversing two perpendicular arms of an optical arrangement a little like that of a Michelson interferometer — a beam-splitter at the centre, directing light in two perpendicular directions, reflecting mirrors at the ends of the two arms and a light intensity meter for measuring phase differences when the two beams have been recombined by the beam splitter.

It is just about ten years that Ronald W. P. Drever, still very much a Scot, was explaining all this in a cramped (and not particularly clean) basement in the physics department of the University of Glasgow. He had then built one leg of an interferometer and was hoping to persuade the British grant agency, the Science and Engineering Research Council, to let him

have a second, longer, leg. Now at Caltech, he is a proud acolyte of Caltech's two-legged instrument, each leg of which is 40-m long. The dream is to scale up by a factor of 100, making each arm 4-km long.

Originally, Caltech and MIT were developing two different optical systems. But for the time being at least, they have combined forces. The sensitivity of an interferometer of the kind being planned in the measurement of the length of the respective legs of the equipment is evidently a direct function of the distance travelled, whence the quest for longer legs. But the same effect can be achieved by making the same light beam traverse each leg repeatedly, provided that the total time for which light is resident there does not exceed that in which there are significant variations of the forces due to gravitational waves.

At Caltech, the existing equipment has been built as if each leg were a Fabry-Perot resonant cavity, which is the agreed basis of the design that will be put forward next year. The idea is to stop the far end of each tunnel with a fully reflecting mirror, and the near end with a nearly matching partly transparent mirror. MIT's design, on ice for the time being, hangs on a trick in geometrical optics whereby rays enter each leg through an off-centre hole in an enclosing mirror, and are reflected repeatedly between one end of each leg and the other until, after a predictable number of double journeys, they re-emerge through the aperture. In a 4-km leg, even 10,000 internal reflections would occupy only about a second.

The immediate goal is to build the pair of 4-km instruments at a sensitivity of 10^{-20} , whereupon two things should be possible: a realistic test of some of the predictions of the kinds of astrophysical phenomena generating detectable gravity waves, and the continued improvement of sensitivity. The obvious difficulty is that improvements are more difficult as the sensitivity is increased. It may be a long time before NSF has to finance a still more ambitious venture. Indeed, the next step will preferably carry the project outside the United States. With only two detectors, unambiguous measures of direction will not be possible. But with a third station, in Europe perhaps, triangulation would be feasible, so that the direction of radiating objects in the sky could be guessed at. That would enormously enhance the project.

John Maddox

Copulation genetics

When incest is not best

Paul H. Harvey and Andrew F. Read

PAUL Greenwood, who spent much of his professional career attempting to unravel the role of inbreeding avoidance in the way animal populations are structured^{1,2}, died this summer. One of his favourite quotations³ was the advice that "you should make a point of trying every experience once, excepting incest and folk-dancing". Yet it is clear that there are circumstances under which, in an evolutionary sense, it would pay to engage in incest^{4,5}. The advantage to a female of mating with a close relative is that she will be helping to propagate additional copies of her own genes if she can provide her relative with more offspring than he would otherwise father. The disadvantage is that the offspring from incestuous matings are typically less fit than those from outbred matings. Natural selection will favour incest avoidance when the costs of incest outweigh the benefits. But what are the costs? As May bemoaned in a News and Views article some years ago⁶, we sorely need a larger body of systematically compiled information on the biological costs of inbreeding. A recent study⁷ by Ralls, Ballou and Templeton provides such data from pedigrees of 40 captive mammal populations belonging to 38 species.

How should the costs of inbreeding be measured? The standard technique was developed by Morton, Crow and Muller⁸, who asked how the rate of juvenile survival decreases with increasing degrees of inbreeding. They estimated A and B from the equation $S = e^{-(A + BF)}$, where S is the proportion of young surviving to some standard age, A is a measure of death resulting from environmental causes and genetic damage in a randomly mating population, F is the inbreeding coefficient (the probability that two alleles present at a given locus are identical by descent), which can be calculated from

pedigrees⁹, and B is a measure of the rate at which survival decreases with increasing inbreeding.

Inbreeding depression is commonly thought to result from the presence, in a single individual, of allelic pairs of recessive deleterious genes which are rare in the population as a whole. Such deleterious recessives are usually masked from expression in outbred individuals but are likely to be inherited in double dose from

offspring mortality following parent-offspring or full sibling mating, whereas golden lion tamarins (*Leontopithecus rosalia rosalia*) have about a 42 per cent increase and brown lemurs (*Lemur fulvus*) a 90 per cent increase.

These costs of inbreeding calculated by Ralls *et al.* are likely to be underestimates, as they do not take into account early embryonic deaths or the fact that many of the young that survived in zoos would have died in the wild. Furthermore, inbred young may well raise fewer offspring, and they could be particularly susceptible to viral infection¹¹. The loss of genetic diversity following inbreeding is also likely to bring additional costs in the long term. There now exist several well-

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Fig. 2 Victim of a bottleneck — a female cheetah with six-week-old young, Kenya.

parents that are close relatives. The action of many such deleterious recessive genes is often estimated in terms of lethal gene equivalents. For example, if an individual contains 20 such recessive genes, each giving a 10 per cent reduction in fitness when expressed as a homozygote, it would be said to contain two recessive lethal equivalents. The value of B above provides a reasonable estimate of the number of recessive lethal equivalents per gamete¹⁰; thus twice that value gives the number per individual. The estimated costs of inbreeding for matings between relatives of different degree are now calculated by Ralls *et al.*⁷ from the estimated value of B using a formula modified from ref. 6. These costs express the additional juvenile mortality expected from an inbred mating over an outbred mating.

In 36 of the 40 populations examined by Ralls *et al.*, juvenile mortality increases with an increasing level of inbreeding. The average cost of a parent-offspring or full sibling mating is 0.33 — mortality is increased by 33 per cent as a consequence of mating with a relative (see Fig. 1) — and the average number of lethal gene equivalents is estimated as 4.6 per individual. One striking result is the wide variation in the costs of inbreeding among species: several exhibit negligible decreases in fitness with increasing levels of inbreeding, whereas others show catastrophic fitness loss. Sumatran tigers (*Panthera tigris sumatrae*), for example, exhibit a barely detectable 0.3 per cent increase in

studied examples of increased vulnerability to infectious disease which has been associated with loss of genetic diversity following a marked reduction in population size (a bottleneck) and inbreeding¹². Perhaps one of the best-known examples is the outbreak of feline infectious peritonitis which decimated several cheetah (*Acyinonyx jubatus jubatus*) populations, a species with very restricted allozyme diversity which is consistent with inbreeding following a bottleneck (Fig. 2). The maintenance of genetic diversity by maximum outcrossing is now an important aim in the management of both wild and zoo populations^{13,14}.

What do the costs of inbreeding reported by Ralls *et al.*⁷ reveal about how females should balance the genetic benefits of mating with close relatives against the fitness costs of inbreeding depression? The average reduction in survivorship of 33 per cent is tantalizing, as it is exactly the figure which one model⁵ suggests as the threshold value beyond which females in polygynous species should not mate with their fathers, brothers or sons, which we shall call incestuous matings. The argument runs as follows. The coefficient of relationship between a mother and the offspring of an incestuous mating is $\frac{3}{4}$ compared with $\frac{1}{2}$ from an outbred mating. But the added benefit to the female must be devalued by the cost of inbreeding. If the reduction in fitness of the incestuous offspring is d , it will pay a female to produce the offspring of her

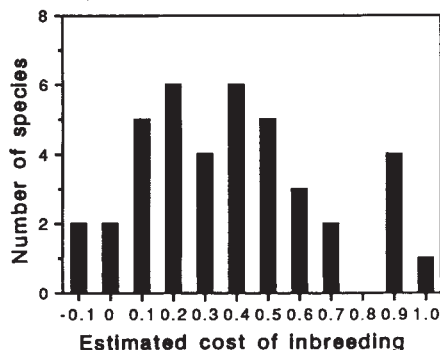


Fig. 1 Estimated costs of producing inbred young, expressed as a proportionate decrease in survivorship, from matings between parents and offspring or full siblings for 40 mammal populations representing 38 species. (From ref. 7.)

father, brother or son rather than a random male in the population if $(1-d)^{3/4} > 1/2$, or $d < 1/3$.

Given that the reported estimates of inbreeding costs are probably too low, it might therefore be expected that matings between close relatives should be rare in natural populations, and that mechanisms of inbreeding avoidance should exist. Recent surveys demonstrate that such matings are indeed rare, and there is scarce but mounting evidence of inbreeding avoidance^{15,16}.

What is the reason for the variation in the costs of inbreeding found among mammals? Because there are no obvious correlates of this variation, it is still not possible to predict the likely costs of inbreeding for any particular species. The previous genetic structure of populations may underlie part of the variation. If a population survives the initial inbreeding depression following a bottleneck, and the deleterious recessive alleles are eliminated in the first few generations, for example, then subsequent levels of inbreeding depression can be considerably lower. Such a situation is believed to have occurred in cheetahs. But the effects of inbreeding depression are not expected to be eliminated in such populations. Indeed, outcrossing in regularly self-fertilizing populations can result in a significant fitness gain¹⁷.

The widespread and usually high costs associated with inbreeding support the view¹⁸ that inbreeding avoidance should be one aim in developing zoo-breeding programmes. But demographic factors may be more important than genetic factors in determining the fate of small natural populations¹⁹. An example is the case¹⁹ of the little spotted owl (*Strix caurina occidentalis*)—a territorial forest-living bird found in north-west North

America—whose future is threatened as its habitat is destroyed by the forestry industry. The original management plan developed by the US Forest Service was designed to protect the territories of about 500 birds on the assumption that this would maintain sufficient genetic diversity in the owl population.

But, using demographic models, Lande argues^{19,20} that suitable habitat is too widely dispersed to prevent the loss of

local populations and eventual species extinction. Despite the undoubted value of estimates of inbreeding depression for understanding the rules of mate choice, there are other crucial variables in the calculations that should be used to provide guidelines for successful management of natural populations. □

Paul H. Harvey and Andrew F. Read are in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

Microelectronics

Fast switching with novel diodes

Serge Luryi

FOR most two-terminal electronic elements (diodes), increasing the voltage applied across them increases the current that they pass, although not necessarily in a linear (ohmic) fashion. But certain semiconductor diodes exhibit a 'negative differential resistance' effect, in which the current decreases with increasing bias voltage. Placed in a resonant circuit and biased into a suitable voltage range such a diode amplifies any circuit oscillations, rather than damping them as does a normal resistance. The classic example of such an element is the Esaki tunnel diode¹, used as a fast, low-noise microwave generator and amplifier. Resonant tunnelling quantum-well diodes are similar devices that promise to operate at frequencies as high as 10^{12} hertz. J. F. Whitaker, T. C. L. G. Sollner *et al.*² have used such a diode in a switching circuit and show that it has a rise time of only 2 picoseconds—the fastest switching event yet observed for an electronic device.

The double-barrier, or quantum-well, diode was first proposed as an electron interferometer analogous to the optical Fabry-Perot étalon. In early proposals^{3,4}, workers considered a semiconductor-metal-semiconductor structure in which the central metal layer would be a resonator for the quantum-mechanical de Broglie waves of the electrons that tunnel from one semiconductor layer (emitter) into the other (collector). At certain energies of incident electrons, the amplitude of the de Broglie waves in the resonator builds up to the extent that these waves leaking in both directions cancel the reflected waves and enhance the transmitted ones.

Thus in the absence of scattering, a system of two identical barriers is completely transparent for electrons entering at a resonant energy and the total transmission coefficients plotted against the incident energy have several sharp maxima. With scattering, the electron wavefunction need not be coherent across the entire double-barrier structure and the electron transport occurs in two sequential tunnelling steps. Negative differential

resistance (NDR) is the property of the first of these steps, involving the tunnelling from a three-dimensional electronic system in the emitter into a two-dimensional system of states in the narrow resonator. The subsequent tunnelling into the collector is an uncorrelated event.

Practical resonators followed the pioneering work of Tsu and Esaki⁵, who showed that an infinite heterojunction superlattice—comprising a periodic sequence of narrow layers of alternate semiconductor materials—should have a state of negative conductivity (and would also sustain oscillations induced by an electric field). In subsequent work⁶ on

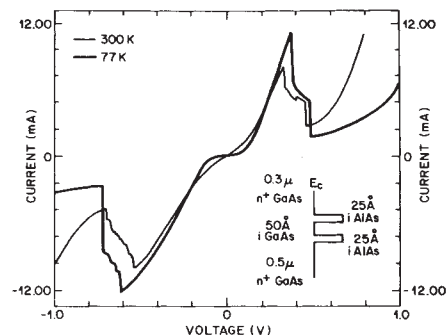


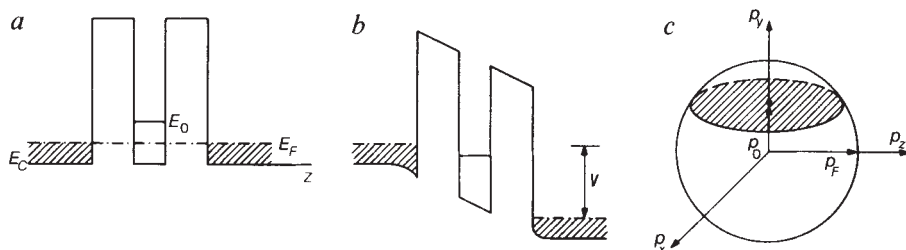
Fig. 1 Current-voltage characteristics of a typical resonant-tunnelling diode (after ref. 11). The structure contains an undoped 5 nm-thick GaAs quantum well clad by two undoped 2.5 nm-thick AlAs barrier layers, as illustrated in the insert. Thermal effects (lattice vibrations) lead to a finite current between the peaks.

finite superlattices, the same authors showed that NDR is possible in double-barrier quantum-well structures, a prediction confirmed experimentally with GaAs/Al_{0.3}Ga_{0.7}As/GaAs double-barrier diodes⁷ and GaAs/AlAs superlattices.

Sollner and co-workers later reported^{8,9} microwave activity in quantum-well diodes with clear NDR characteristics at 77 K. They showed that the diode could operate as a detector and mixer of far-infrared radiation at 2.5 terahertz (10^{12} Hz), and obtained active oscillations at frequencies up to 18 gigahertz from a diode mounted in a resonant cavity.

- Greenwood, P.J. *Anim. Behav.* **28**, 1140–1162 (1980).
- Greenwood, P.J., Harvey, P.H. & Perrins, C.M. *Nature* **271**, 52–54 (1978).
- Bax, A. *Farewell my Youth* (Longman, London, 1943).
- Wilson, F.O. in *Theoretical Ecology* (ed. May, R.M.) 205–218 (Blackwell, Oxford 1976).
- Smith, R.H. *Heredity* **43**, 205–211 (1979).
- May, R.M. *Nature* **279**, 192–194 (1979).
- Ralls, K., Ballou, J.D. & Templeton, A. *Conserv. Biol.* **2**, 185–193 (1988).
- Morton, N.E., Crow, J.F. & Muller, H.J. *Proc. natn. Acad. Sci. U.S.A.* **42**, 855–863 (1955).
- Ballou, J. in *Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations* (eds Schonewald-Cox, C.M. *et al.*) 509–520 (Cummings, Menlo Park, 1983).
- Cavalli-Sforza, L.L. & Bodmer, W.F. *The Genetics of Human Populations* (Freeman, San Francisco, 1971).
- O'Brien, S.J. *Science* **227**, 1428–1434 (1985).
- O'Brien, S.J. & Evermann, J.F. *Trends Ecol. Evol.* **3**, 254–259 (1988).
- Ralls, K. & Ballou, J.D. *Zoo Biol.* **5**, 81–238 (1986).
- Read, A.F. & Harvey, P.H. *Nature* **322**, 408–410 (1986).
- Ralls, K., Harvey, P.H. & Lyles, A.M. *Conservation Biology: The Science of Scarcity and Diversity* (ed. Soulé, M.) 164–184 (Sinauer, Sunderland, Massachusetts, 1986).
- Harvey, P.H. & Ralls, K. *Nature* **320**, 575–576 (1986).
- Charlesworth, D. & Charlesworth, B. *A. Rev. Ecol. Syst.* **18**, 237–268 (1987).
- Ralls, K. & Ballou, J. in *Genetics and Conservation: A Reference for Managing Wild Animal and Plant Populations* (eds Schonewald-Cox, C.M. *et al.*) 164–184 (Cummings, Menlo Park, 1983).
- Lande, R. *Science* **241**, 1455–1459 (1988).
- Lande, R. *Am. Nat.* **130**, 624–635 (1987).

Fig. 2 Operation of a double-barrier resonant-tunnelling diode. *a*, The electron energy diagram without biasing. *b*, With an applied bias V , for which the energy of certain electrons in the emitter matches unoccupied levels of the lowest sub-band E_0 in the quantum well. *c*, The 'Fermi sphere' of occupied states for the three-dimensional electron 'gas' in the emitter — the locus of points in the momentum (p) space with $p_x^2 + p_y^2 + p_z^2 \leq p_F^2 = 2mE_F$ (m is the electron mass). Conservation of the lateral momentum during tunnelling requires that only those emitter electrons whose momenta lie on a disk $p_z = p_0$ (shaded) can tunnel into the quantum well. Here, $p_0^2/2m$ is the energy separation between E_0 and the bottom edge E_c of the conduction band in the emitter. The number of emitter electrons clearly increases as p_0 decreases. In an ideal double-barrier diode at 0 K, resonant tunnelling occurs in a voltage range in which the disk moves down from the pole to the equatorial plane of the emitter Fermi sphere. At higher V (when $p_0^2 < 0$) resonant electrons no longer exist, which results in a sharp drop in the current.



Although the basic physical phenomena of resonant tunnelling were previously anticipated, realization of their full potential had to wait until recent improvements in epitaxial techniques. Indeed, materials are now so good that NDR can be readily observed at room temperature (Fig. 1).

Negative differential resistance arises from energy and momentum conservation

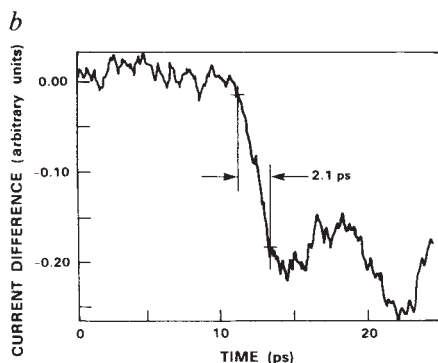
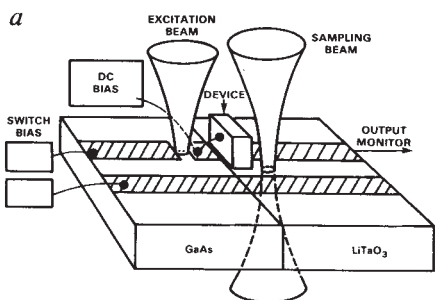


Fig. 3 The experiment of Whitaker *et al.*². *a*, Schematic of apparatus. The optical excitation beam closes a photoconductive switch with an optical pulse lasting only 80 femtoseconds. The sampling beam measures the electric field in the transmission line after the electrical pulse has passed through the resonant-tunnelling diode. The sampling beam is synchronized with the probe beam, so the temporal behaviour of the propagating electrical pulse can be obtained with a resolution of less than a picosecond. *b*, The switching of a resonant-tunnelling diode from its high-current state to its low-current state. The electrical pulse from the excitation beam shown in (*a*) drives the diode above its peak current to a position near the valley current. The circuit allows the bias point to return to its original value so that the next pulse duplicates the action of the previous one. (Courtesy of T.C.L.G. Sollner.)

and from the electron confinement in the quantum-well resonator (Fig. 2). The confinement distinguishes states of motion along the z -direction (perpendicular to the interfaces) inside the well from those outside. Those inside form a ladder of discrete levels, $E_0, E_1, \dots$, whereas those outside form a continuum of energy states. (The unconfined motion parallel to the interfaces both inside and outside the well, contributing an energy continuum, has only a secondary effect in the process.) Resonant tunnelling into the well occurs only for electrons whose 'forward' kinetic energy in the emitter matches a level in the well (Fig. 2). Biasing effectively lowers the states inside the well relative to the emitter, changing the emitter population that participates in tunnelling.

Tunnelling starts as biasing brings a well level down to the maximum (Fermi) energy E_F of electrons in the emitter. Further biasing increases the number of electrons participating (Fig. 2c) so that the conductance increases — the resistance decreases. Eventually, further biasing removes the well level below the minimum energy E_c of conducting electrons and tunnelling ceases. (Imperfections and lattice vibrations at finite temperatures broaden the cut-off and permit a small residual current.) The process is repeated as further biasing brings the next well level down to the Fermi energy.

Biased into one of its NDR regions, the double-barrier diode amplifies oscillations inherent to any resonant circuit with which it is connected. Such spontaneous oscillations often obscure the measurement of the diode's stable current-voltage characteristic. But when mounted in an appropriate cavity, the diode can generate microwaves, wherein lies its main utility. The critical parameter for this use of resonant-tunnelling diodes is the maximum frequency which can be obtained. This frequency is fundamentally limited by the inverse lifetime of electronic states in the quantum-well resonator.

If scattering is not important, the lifetime τ is controlled by the tunnelling processes in and out of the quantum well. In this case, the diode operation is described by the Fabry-Perot picture. For relatively

thick barriers (more than 5 nanometers) one can have an extremely sharp resonance (long τ), but this would give a low limit frequency. The double-barrier diodes studied by Whitaker *et al.*² were implemented in a GaAs/AlAs heterostructure with extremely narrow barriers (1.5 nanometres), grown by molecular beam epitaxy. The estimated tunnelling lifetime was of the order of one picosecond which translates into a maximum oscillation frequency of about 200 gigahertz. E.R. Brown, W. D. Goodhue and Sollner recently reported¹⁰ measurements of small-signal fundamental oscillations up to this frequency in these diodes.

The large-signal switching experiments reported by Whitaker *et al.*² have more in common with the operation of logic elements. The authors performed a real-time measurement by an electro-optic sampling technique, measuring the rise time of switching in a circuit containing the double-barrier diode loaded on a small positive resistor. Because of the NDR characteristic of the diode, such a circuit exhibits bistability in a certain range of the applied bias. If the bias condition changes, the circuit switches. The tricky problem, solved by Whitaker *et al.* (Fig. 3), is to change the bias and sample the circuit response at picosecond speeds. The switching is accomplished in about 2 picoseconds, demonstrating that quantum-well diodes can have a response time comparable to that of fastest all-optical logic elements. Thus these devices could become the basis of future high-speed computers. □

1. Esaki, L. *Phys. Rev.* **109**, 603 (1958).
2. Whitaker, J.F., Mourou, G.A., Sollner, T.C.L.G. & Goodhue, W.D. *Appl. Phys. Lett.* **53**, 385-387 (1988).
3. Davis, R.H. & Hosack, H.H. *J. appl. Phys.* **34**, 864 (1963).
4. Iogansen, L.V. *JETP* **18**, 146 (1963).
5. Esaki, L. & Tsu, R. *IBM J. Res. Dev.* **14**, 61 (1970).
6. Tsu, R. & Esaki, L. *Appl. Phys. Lett.* **22**, 562 (1973).
7. Chang, L.L., Esaki, L. & Tsu, R. *Appl. Phys. Lett.* **24**, 593 (1974).
8. Sollner, T.C.L.G., Goodhue, W.D., Tannenwald, P.E., Parker, C.D. & Peck, D.D. *Appl. Phys. Lett.* **43**, 588 (1983).
9. Sollner, T.C.L.G., Tannenwald, P.E., Peck, D.D. & Goodhue, W.D. *Appl. Phys. Lett.* **45**, 1319 (1984).
10. Brown, E.R., Goodhue, W.D. & Sollner, T.C.L.G. *J. appl. Phys.* **64**, 1519 (1988).
11. Morkoç, H., Chen, J., Reddy, U.K., Henderson, T. & Luryi, S. *Appl. Phys. Lett.* **49**, 70-72 (1986).

Serge Luryi is at AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA.

Neurobiology

Are single retinal neurons both excitatory and inhibitory?

Robert F. Miller

ONE of the most intriguing findings presented at a recent meeting* emerged from studies of 'starburst' amacrine cells (R. Masland and D. O'Malley, Harvard Medical School). These cholinergic amacrine cells were first described in the inner plexiform layer of the rabbit retina (see Fig. 1), and have since been identified in several other species. These cells synthesize and release acetylcholine (ACh), which provides excitatory input to many types of ganglion cells. Three laboratories (Brecha, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6187; 1988; Vaney, D. I. & Young, H. M. *Brain Res.* **438**, 369; 1988 and Kosaka, T. *et al. Expl Brain Res.* **70**, 605; 1988) have independently provided evidence which suggests that, in addition to their cholinergic function, starburst amacrine cells synthesize, store and release GABA (γ -aminobutyric acid), a powerful inhibitory neurotransmitter.

Although co-localization of more than one neurotransmitter in a single neuron has been established at numerous sites in the nervous system (including several subtypes of amacrine cells in the retina), this seems to be the first evidence for co-localization of two 'fast-acting' neurotransmitters that have opposite modes of action on postsynaptic cells (excitation and inhibition). Co-release of both transmitters could cancel the actions of each at postsynaptic sites, because ganglion cells are known to be sensitive to the application of both GABA and ACh. But Masland's and O'Malley's experiments on the release of radio-labelled GABA and ACh from rabbit retina show that ACh release is modulated by light stimulation and is calcium-dependent, consistent with the idea that it occurs through discharge of synaptic vesicles (see Fig. 2). In contrast, GABA efflux is modulated neither by light nor by changes in external calcium. Addition of KCl to the bathing medium produces a large increase in GABA and in ACh release, demonstrating that GABA release from the tissue can be accelerated, presumably as a result of the depolarizing action of potassium ions.

Masland and O'Malley also used antibody-labelled beads to precipitate synaptic vesicles, showing that these vesicles contain a high fraction of the retinal stores of ACh, but no detectable GABA activity. Thus, they argue that GABA release is mediated through a carrier system, such as that proposed for GABA efflux from horizontal cells of the amphibian retina (E. Schwartz, University of Chicago), whereas ACh release is mediated through the better-known vesicular exocytotic pathway.

One obvious question about starburst amacrine cell function arising from these data relates to the mechanism of GABA release and its relative insensitivity to light stimulation. In amphibians it is known that light-evoked GABAergic synaptic responses are present in some ganglion cells of the retina. How can there be a large basal release of GABA, sensitivity of the release to external K^+ , but relative insensitivity to light stimulation? Perhaps ACh and GABA release are modulated by light but ACh release is highly amplified.

A subclass of retinal ganglion cells are directionally selective, responding with impulse activity to moving targets in one direction (preferred) but not in the other (null). These cells are known to receive strong excitatory cholinergic input, and

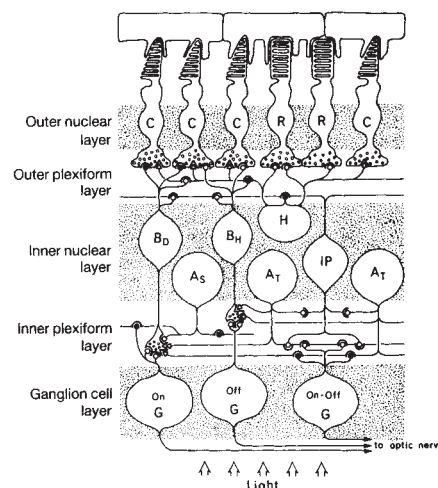


Fig. 1 The basic cellular and synaptic organization of the vertebrate retina. Excitatory (open circles), inhibitory (closed circles) and reciprocal (triangles) synapses are shown in each plexiform layer. Two types of bipolar cells are shown, those that depolarize to spot illumination (B_D), and those that hyperpolarize to spot illumination (B_H). A_S , amacrine cells that respond to light with sustained potentials, A_T , those that respond with transient potentials. The cholinergic amacrine cells are not illustrated but consist of complementary on and off cells, one of which (on) has its cell body in the ganglion cell layer while the second (off) type has its cell body in the amacrine cell layer. (From *Encyclopedia of Neuroscience*, Vol. II, ed. Adelman, G.; Birkhäuser, Boston, 1987).

pharmacological evidence suggests that GABA provides the inhibitory input in the null direction. Is a single cell type, the cholinergic amacrine cell, responsible for both excitation in the preferred direction and inhibition in the null direction? A descriptive model addressing this fascinating possibility was presented by D. I. Vaney (National Vision Research Institute, Australia).

Masland and O'Malley's work raises the possibility that the organization of cells at the sub-cellular level is important in the functional organization of synaptic pathways. What are the mechanisms of GABA release and how are they coupled to the membrane potential? Are the release sites for GABA co-distributed with those of ACh or are there regional differences? How are the postsynaptic GABA and ACh receptors distributed? Can the different mechanisms of release help to account for specialized synaptic functions such as directional selectivity? Answering these questions will require substantial technical advances. Furthermore, GABA release in the retina of the frog *Xenopus* (M. Neal, St Thomas's Hospital, London) has a substantial calcium-

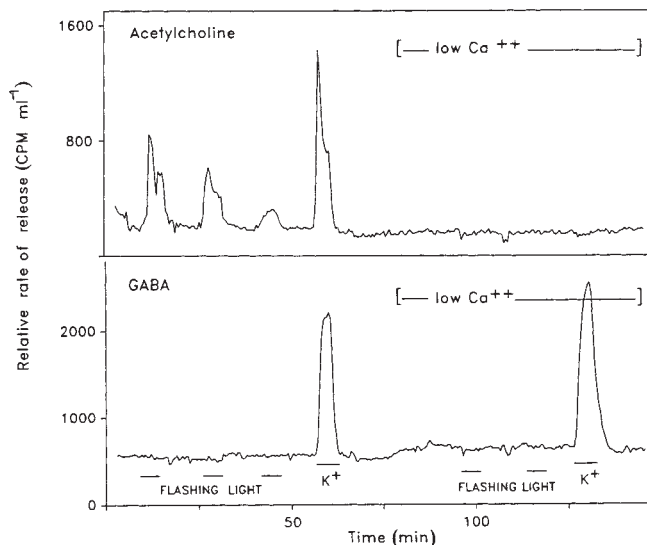


Fig. 2 Differential control of the releases of ACh (top) and GABA (bottom) from a rabbit retina. When the retina was stimulated by flashing light, the cells released ACh, but there was no measurable increase in their release of GABA. Raising the extracellular concentration of K^+ to 30 mM caused a release of both acetylcholine and GABA. Raising the concentration of K^+ in a low Ca^{2+} medium (to prevent the Ca^{2+} influx that initiates vesicle exocytosis) caused a release of GABA but no release of ACh. This suggests that ACh is released by vesicle exocytosis and GABA is released via a carrier.

* *The Neurobiology of the Inner Retina*. NATO Advanced Study Workshop, Oldenburg, West Germany, 8-12 August 1988. Proceedings to be published by Springer.

sensitive component, thus emphasizing that species differences in GABA release mechanisms may limit the generalities that can be deduced from studies in a single animal.

The meeting served two important functions. First, it was a reminder of the complexity of the inner plexiform layer of the vertebrate retina; recent findings in some species suggest that as many as 40–50 different morphological subtypes of amacrine cells may be present, with about two dozen known neuroactive substances. Second, although the amacrine-cell layer is rich in its diversity of neuron forms and neurotransmitter species, the main inputs

that convey the receptive-field properties of retinal ganglion cells are derived from a small subset of transmitter types, but a large subset of the neuronal population. One estimate suggests that 90 per cent of retinal synapses can be accounted for by glutamate (photoreceptors and bipolars), GABA and glycine (amacrine cells). But the remaining 10 per cent of synapses, whose functions remain to be elucidated, must be involved in the long-neglected processes of homeostasis. □

Robert F. Miller is in the Department of Physiology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA.

Atomic physics

Order and chaos in strong fields

D. Richards

ATOMS are generally remarkably stable objects: viewed as miniature solar systems, as in the Bohr–Rutherford model, their constituent electrons execute countless unchanging orbits. This stability is well described by quantum mechanics. But extreme conditions are now available in the laboratory in which atoms are accurately described by classical physics. These developments and the parallel improvements in theory were the subject of a recent meeting*. In particular, the prediction and observation of chaotic behaviour of atoms in strong external fields confirm the significance of these developments.

Under normal conditions, the behaviour of atoms is dominated by the central electrostatic pull of the nucleus, with corrections for electron correlations, magnetic interactions and other effects. The effect which an external force can have on an atom depends on, amongst other variables, the ratio of the force exerted on the electrons by the external field to the forces binding the electrons to the nucleus.

For example, standard modern pulsed lasers, having peak intensities in the range 10^{12} – 10^{14} W cm⁻², produce peak electric fields of 10^7 – 10^8 V cm⁻¹, although peak fields of the order of 10^{11} V cm⁻¹ have been achieved (C.K. Rhodes, University of Chicago). These fields should be compared to typical atomic field strengths of around 5×10^9 V cm⁻¹. The limited magnitude of the typical peak fields limits their effect on the behaviour of the atoms and the conventional methods of atomic physics explain most of the observed processes, such as multiphoton ionization (A. Szöke, Lawrence Livermore Laboratory). Some of this agreement probably arises from poor experimental resolutions (T.J. McIlrath, University of Maryland).

Laboratory magnetic fields of up to 10 tesla (T) are routinely available, but these are also very small compared with atomic fields of 2.4×10^5 T. But on white dwarfs and neutron stars fields of 10^2 – 10^5 and 10^5 – 10^9 T, respectively, are found. Understanding the effects that these gigantic fields have on the atomic spectra helps understand the physics of these stars: in particular, the stellar magnetic field strength can be determined more precisely (G. Wunner, University of Tübingen).

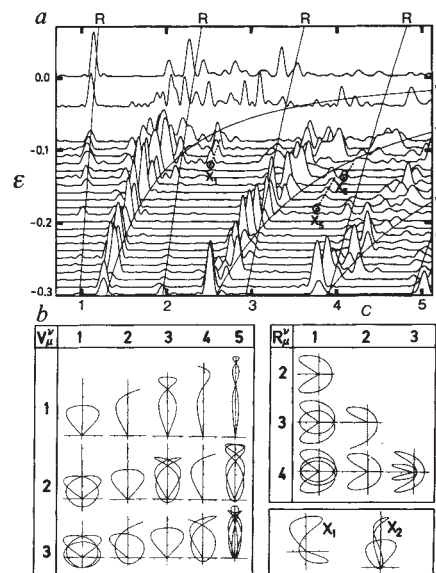
The constraint imposed by relatively weak laboratory fields is removed by developments in experimental techniques which permit experiments on Rydberg atoms. These have a single electron in a highly excited state and are characterized by the principal quantum number, n , which can vary from 5 up to 300 in current experiments. The excited electron is very weakly bound, the binding force varying as n^{-4} and the characteristic magnetic force varying as n^{-3} . Thus by varying n and the applied laboratory field the important ratio of field strengths can be varied over a wide range.

Paradoxically, the experiments (J.E. Bayfield, University of Pittsburgh; P.M. Koch, State University of New York at Stony Brook; K.H. Welge, University of Bielefeld) on these delicate atoms are more accurate than many equivalent experiments on ground state atoms. In Koch's microwave experiments, the field strengths are known to better than ± 5 per cent and the field envelope is known very accurately. These experiments show such systems to have very complex dynamics.

Increasing the principal quantum number n has, however, a more profound effect. For large values of n , atoms respond like classical atoms whose motion is described by the nonlinear equations of Hamilton, rather than the linear wave equation of Schrödinger conventionally

used in atomic physics. For large values of n , the wavefunction comprises many constituent waves which interfere constructively along classical orbits and destructively away from these orbits. Thus the electron is constrained to its classical orbit, its wave-like behaviour suppressed, and it tends to behave deterministically. Similar effects are seen in light caustics which are concentrations of light along rays, the equivalent of classical orbits. Because the solutions of nonlinear and linear equations are qualitatively different, new effects should be seen as n increases. Indeed, the complexities of the solutions of nonlinear equations and the subtleties of the interference effects mean that the large n limit is very intricate.

For a classical atom in a strong field, most orbits are chaotic and wander indiscriminately over the whole available energy shell. In this chaotic sea there are infinitely many periodic orbits, some stable and some unstable, although they comprise only a fraction of all possible trajectories. In this limit, one expects that observed spectra will comprise a quasi-continuum background, due to the chaotic orbits, superimposed with sharp peaks



Resonances in the spectrum of the chaotic even-parity ($m=0$) hydrogen atom. *a*, Individual spectra at different scaled energy ($\epsilon = EB^{-2/3}$) are overlaid, with the intensity adjusted for clarity. Abscissa, scaled action ($C = nB^{1/3}$); E and B are the electric and magnetic field strengths, n the principal quantum number. Superposed curves highlight the predicted classical resonances, classed as vibrational (one-dimensional; V), rotational (two-dimensional; R) and exotic (dotted curves marked X). Subsets of vibrational curves (not shown) fan out, or bifurcate, to the left of the root curves shown, and also coincide with observed resonances. Chaotic trajectories are not observed directly, but contribute to the background signal. *b*, Calculated classical periodic orbits corresponding to the observed spectral peaks, plotted in the cylindrical coordinates ρ, z . (From Holle, A. *et al. Phys. Rev. Lett.* 61, 161–164; 1988.)

* *Atoms in Strong Fields*, Grainau, West Germany, 5–9 September 1988.

due to the periodic orbits. Further, the energy-time uncertainty principle, $\Delta E \sim \hbar/t$ ($\hbar$ is Planck's constant) suggests that orbits of period t will produce structure of width ΔE and conversely experiments of resolution ΔE will reveal the effects of classical orbits of period t . These predictions, based on classical and semiclassical theories (J.B. Delos, The College of William and Mary; E. Heller, University of Washington; W.P. Reinhardt, University of Pennsylvania), have been confirmed in various model calculations. In some cases (Reinhardt) one can Fourier analyse experimental spectra to infer the periods of relevant classical orbits.

The most convincing demonstration, however, is the comparison of the absorption spectra of excited hydrogen in a strong magnetic field (Welge) in which is seen an almost perfect correlation between the observed structure with the relevant computed classical orbits, some of which have quite exotic shapes (see figure). Even more spectacular is the observation in this atomic system of features which appear to be caused by period doubling in the classical orbits, a diagnostic signature of classical chaos.

The importance of the relatively few periodic orbits was emphasized by many. The spectra of excited molecules usually contain far too many lines to be classified using conventional methods; low-resolution spectroscopy usually enhances the main features due to periodic orbits, which can be used to construct a relatively small, relevant basis for a purely quantum

mechanical calculation that can accurately describe these features (H. Taylor, University of Southern California). This technique should provide a sensitive test of theoretical potential energy surfaces and help understand the spectra of complex systems. The effect of periodic orbits is also seen in microwave ionization measurements of excited hydrogen (Koch), although in this case there are clear quantum effects associated with these orbits which are not fully understood.

The chaotic classical trajectories are associated with the background sea of spectral lines, which are amenable to statistical analysis (O. Bohigas, Institute Physique Nucléaire, Orsay) as are the intensities of these lines (Wunner). This approach elegantly connects the dynamics of few-body systems with the statistical methods used in the many-body theories of the nucleus.

Classical dynamics, however, does not always work. In some cases the solution of Schrödinger's equation cannot follow the classical trajectories of Hamilton's equations because interference between the constituent waves is too weak. One example is the high-frequency ionization of excited hydrogen (Bayfield; Koch) showing a clear breakdown of classical dynamics as predicted (G. Casati, University of Milan). This disagreement is now accepted, but the detailed reasons are still hotly debated. □

D. Richards is in the Faculty of Mathematics, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK.

Nuclear physics

High-density equation of state

G. E. Brown

JUST as the state of ordinary matter depends on its density and energy content, so does the state of nuclear matter. But experiments to determine the equation of state at high density, which bears closely on the forces that bind nuclear matter, are almost impossible to perform. On page 560 of this issue¹, J. L. Friedman *et al.* revive the idea^{2,3} of using neutron stars, composed entirely of dense neutron matter, as cosmic nuclear laboratories. They argue that, because of the apparent absence of pulsars (spinning neutron stars) with spin periods of less than 1.55 milliseconds, nuclear matter must be highly incompressible, or 'stiff'. Their conclusion will invigorate the continuing debate on whether nuclear matter is stiff or 'soft'.

The equation of state (EOS) of nuclear matter, which relates the pressure in the matter to its density, reveals aspects of nuclear interactions, but also has important astrophysical implications. In particu-

lar, the outcome of supernova explosions of exhausted stars depends crucially on the response of nuclear matter to gravitationally induced compression. Conversely, supernovae provide a test of proposed equations of state.

The envelope of a large star comprises the elements from hydrogen up to iron, the lighter elements having burned into the heavier ones throughout the life of the star. It is necessary that these be blown off, because such a cycle is the only known mechanism for releasing to the galaxies the elements from carbon through to iron (and heavier ones formed through more complicated processes).

Calculations show that supernovae, which follow the collapse of large stars⁴, occur only if the EOS of the somewhat neutron-rich matter is soft up to densities around $4\rho_0$, where ρ_0 is the density of ordinary nuclei. (Calculations which allow for several 'flavours', or types, of neutrinos, require an even softer EOS.)

The soft EOS offers little resistance against the initial collapse of the stellar core, so that the collapse is halted only at extremely high nuclear densities. In the subsequent rebound, the compressed core drives out like a piston powering the explosion which blows off the surrounding envelope. If the envelope were not blown off, it would fall back on to the core, which would collapse into a black hole. This happens in our calculations when the EOS is stiffened.

But in this issue¹, Friedman *et al.* suggest that the pulsar PSR1957+20 is spinning near its limiting frequency and conclude that the EOS of its neutron matter is stiff. In this case the star is larger than if the matter is soft, and the material at the equator will be rotating near the maximum velocity it can have without flying off from the star. Thus, more rapidly rotating stars would break apart. This concurs with earlier work on the 35-day periodicity in the X-ray intensity from the object Her X-1, which Pines and co-workers^{2,3} interpret as a freely precessing neutron star. The observed wobble frequency can be explained if the star has a large elastic energy content. This would be the case if the star has a thick crust, which requires the EOS of neutronic matter to be stiff.

One might believe that the EOS of nuclear matter is soft and that of neutron matter, stiff. But if the latter is made too stiff, neutrons (n) simply change the protons (p) through the reaction $n \rightarrow p + e^- + \nu$, with the neutrino ν escaping (e^- is an electron). In this way a thermodynamic equilibrium between protons and neutrons is established. This brings the composition and pressure of the system closer to that of nuclear matter; the additional electrons produce little pressure. Thus, the neutron EOS cannot be made very much stiffer than that for nuclear matter.

It is possible that the EOS is soft for densities ρ_0 – $4\rho_0$ and stiff for higher densities. Stellar collapse, which precedes a supernova, explores only the former range, whereas given a soft EOS in this range, most of the neutron star will settle into higher densities. This is the scheme used by Prakash *et al.*⁵ who begin from an EOS which is extremely soft at low densities but nonetheless succeed in predicting stable neutron stars up to masses of nearly 1.5 solar masses ($M_\odot$).

It is possible that the prompt mechanism described above for supernova explosions is not correct. A delayed mechanism has also been proposed^{6,7}, in which the explosive shock-wave temporarily stagnates until it is repowered by neutrinos from the hot core. But calculations give less than half the required explosion energy.

Stiff equations of state permit stable neutron stars with masses greater than $2 M_\odot$. But because many neutron stars occur in binaries, where they can accrete material from a companion star, it is strange that no-one has seen any of these

massive stars. Accurately measured neutron star masses all fall near $1.4 M_{\odot}$, the two in the neutron star binary PSR1913+16 being the most accurately measured⁸: $1.444 \pm 0.01 M_{\odot}$ and $1.384 \pm 0.01 M_{\odot}$. The mass of 4U0900-40 is quoted⁹ as $1.85 \pm 0.3 M_{\odot}$, but the error is large and this mass could well be less than $1.5 M_{\odot}$. Very soft equations of state stabilize masses only less than $1.5 M_{\odot}$. Maybe none are larger than this because the EOS is indeed soft and they cannot be stabilized.

In 'laboratory' physics, the EOS, expressed as the compression modulus K_0 of nuclear matter, has been determined chiefly from the 'breathing' vibrational mode of ^{208}Pb . From this state, K_0 is determined¹⁰ to be $210 \pm 30 \text{ MeV}$, which is very stiff compared to the value ($K_0 = 120 \text{ MeV}$) used by Prakash *et al.* in their soft-EOS calculation⁵ for neutron stars. But correcting for quantum effects and allowing for the finite size of the nucleus ^{208}Pb , I find¹¹ that for infinite nuclear matter $K_0(\infty) \leq 2/3 K_0(^{208}\text{Pb}) \leq 140 \text{ MeV}$. Even lower values seem to be required¹² to fulfil a sum rule for interactions developed by Landau.

Thus observations do not yet distinguish decisively between the two 'camps' — the stiff and the soft — of those studying the nuclear EOS. But the observation of a pulsar with a period of less than 1.55 milliseconds would negate the argument¹ of Friedman *et al.* in favour of a stiff EOS; or the discovery of a neutron star more massive than $1.5 M_{\odot}$ would destroy my argument for a very soft EOS.

Many-body calculations for supra-nuclear densities predict a rather stiff EOS, with $K_0 < 210 \text{ MeV}$. But it seems that the masses of the scalar and vector mesons — the carriers of the nuclear interaction — decrease with increasing nuclear density. Until recently, these masses were assumed to be constant. Introducing such density-dependent meson masses changes the calculations drastically, permitting a substantial, but as yet uncertain, softening of the EOS. □

1. Friedman, J. L., Imamura, J. N., Durisen, R. H. & Parker, L. *Nature* **336**, 560–562 (1988).
2. Pines, D. in *Proc. 13th Texas Symp. Rel. Astrophys.* (ed. Ulmer, M.P.) 440–466 (World Scientific, Singapore, 1987).
3. Pandharipande, V. R. *et al. Astrophys. J.* **205**, 550–566 (1976).
4. Baron, E., Cooperstein, J. & Kahana, S., *Phys. Rev. Lett.* **55**, 126–129 (1985).
5. Prakash, M., Ainsworth, T. L. & Lattimer, J. M. *Phys. Rev. Lett.* **61**, 2518–2521 (1988).
6. Wilson, J. R. in *Numerical Astrophysics* (eds Centrella, J. *et al.* 422–434 (Jones & Bartlett, Boston, 1985).
7. Bethe, H. A. & Wilson, J. R. *Astrophys. J.* **295**, 14–23 (1985).
8. Weisberg, J. M. & Taylor, J. H. *Phys. Rev. Lett.* **52**, 1348–1350 (1984).
9. Joss, P. C. & Rappaport, S. A. *A. Rev. Astr. Astrophys.* **22**, 537–592 (1984).
10. Blaizot, J. P., Gogny, D. & Grammaticos, B. *Nucl. Phys. A* **265**, 315–336 (1976).
11. Brown, G. E. *Nucl. Phys. A* **488**, 689c–719c (1988).
12. Pines, D., Quader, K. F. & Wambach, J. *Nucl. Phys. A* **477**, 365–398 (1988).

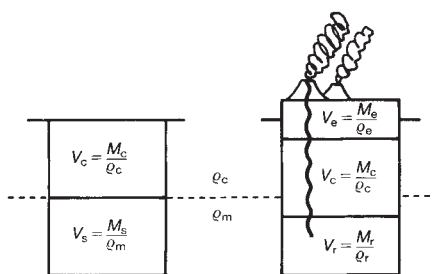
G. E. Brown is in the Institute for Theoretical Physics, State University of New York, Stony Brook, New York 11794, USA.

Martian geology

Making a big impression

Michael H. Carr

THARSIS bulge, discovered on Mars in the early 1970s, is about 8,000 km across, 10 km high at its centre and covers over a quarter of the planet's surface. It clearly has played a significant part in the evolution of the planet. Most of the volcanic activity on Mars, including some of the largest shield volcanoes in the Solar System, has been concentrated there: huge equatorial canyons start at the centre of the bulge and extend down the eastern flank, and most of the large flood features of the planet are around its periphery. The bulge is also at the centre of a vast radial



Tharsis bulge could be the result of volume changes resulting from volcanic activity. (V , volume; M , mass; and ρ , density. Subscripts: c, crust; m, mantle; s, source; e, erupted; and r, residual.) Eruptions remove mass M_e of rock from the mantle, building a block of new material on the surface. The combined volume of the residual and erupted rock is greater than that of the source rock, so the surface becomes raised. (After Finnerty *et al.*)

fracture system that affects half the planet. How did the bulge form and how has it sustained itself throughout martian history? A.A. Finnerty *et al.* now suggest (*J. geophys. Res.* **93**, 10225–10235; 1988) that changes of density that resulted from vertical transport of volcanic magmas caused the bulge.

Theories about the origin of the Tharsis bulge can be classified as either 'uplift' or 'constructional' models. The first class suggest that the martian lithosphere in the Tharsis region has been subject to uplift from below. The uplift can result from isostatic forces attempting to compensate for density anomalies in the mantle and crust, or from dynamic processes such as mantle convection (Wise, D.U., Golombek, M.P. & McGill, G.E. *Icarus* **38**, 456–472; 1979). Constructional models suggest that the bulge is due largely to near-surface accumulation of volcanics (Solomon, S.C. & Head, J.W. *J. geophys. Res.* **87**, 9755–9774; 1982). The new model proposed by Finnerty *et al.* is a variant of these models, in the sense that the bulge is caused by accumulation of near-surface volcanics. Uplift of the

surface occurs only insofar as the near-surface volcanics are intruded below the surface rather than extruded onto it.

There is a large gravity anomaly associated with the Tharsis bulge which N.H. Sleep and R.J. Phillips (*Geophys. Res. Lett.* **6**, 803–806; 1979) showed could have formed isostatically, without the introduction or loss of material from the Tharsis region. They suggested that the gravity field could be explained if anomalously high near-surface densities were compensated by anomalously low mantle densities that extend to depths of at least 300 km. The petrologic model proposed by Finnerty *et al.* is consistent with the isostatic gravity model. The authors suggest that the Tharsis bulge is the result of large-scale volcanism. The model requires no lateral transport of material, only vertical transport of magmas from their source regions, deep within the mantle, towards the surface (see figure).

Volcanism would result in extraction of iron-rich basaltic melts from the mantle, leaving it enriched in magnesium, and hence anomalously buoyant. Deposition of the iron-rich magmas on or near the surface would leave the near-surface materials anomalously dense. Finnerty *et al.* use previous experimental data to estimate the range of mantle densities expected, and the densities expected of the extracted magmas and resulting volcanic rocks. They estimate that 30–50 per cent of the source region was partially melted to build Tharsis, the fraction being a function of the densities of the various crust and upper-mantle components, and conclude that the lithosphere under Tharsis must be at least 300 km thick.

The model is ingenious but still leaves several unanswered questions. Such a lithospheric thickness contrasts markedly with that estimated (20–30 km) from the flexing of the lithosphere under local loads (Comer, R.P., Solomon, S.C. & Head, J.W. *Rev. Geophys.* **23**, 61–92; 1985). What triggered the massive volcanism in Tharsis? Why was the process so sustained in Tharsis, while most of the rest of the planet was left almost unaffected? Fortunately, the theory could soon be tested, for Mars Observer, to be launched in 1992, will greatly improve our knowledge of the planet's gravity and topography, and will also reveal something of the chemistry and mineralogy of the near-surface rocks, about which Finnerty *et al.* make specific predictions. □

Michael H. Carr is in the US Geological Survey, Menlo Park, California 94025, USA.

Plant ecology

Spatial dynamics of self-thinning

Gareth Hughes

FIELD data suggest that density-dependent mortality (self-thinning) in even-aged, single-species plant populations is accompanied by changes in spatial pattern, towards a more regular distribution of surviving plants. A substantial body of such data relates to natural populations of forest trees, where practical difficulties deter study of the development of spatial pattern over time. Instead, changes in spatial pattern are inferred from comparisons of the patterns of otherwise similar populations of different ages¹ or, as in a new study² by Kenkel, from comparisons of the patterns of live and dead members of a single population.

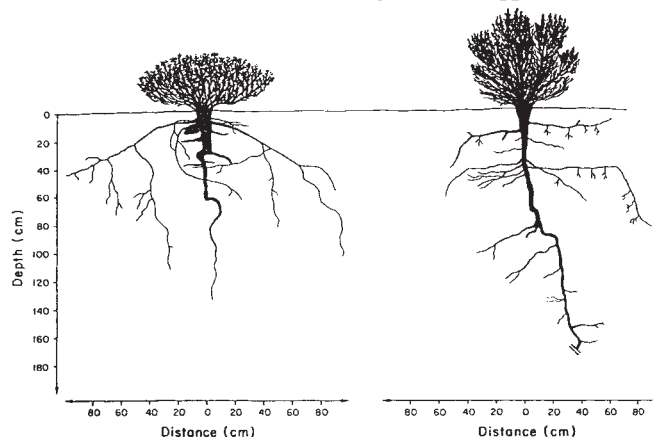
Kenkel used various spatial statistics to analyse pattern in a 65-year-old jack pine (*Pinus banksiana*) population growing in a relatively homogeneous physical environment. The pattern of the initial (live plus dead) population shows no departure from randomness, but when both the dead and the live trees are considered separately, they show non-random patterns. The dead trees are more aggregated than can be accounted for by random mortality, whereas the survivors tend to have a regular pattern.

Spatial inter-relationships between live and dead trees indicate that the development of a regular pattern of live trees cannot be wholly explained by differential mortality in higher density patches in the initial population. Kenkel suggests that these results are consistent with a two-phase model of intraspecific competition: first, competition for soil resources leads to thinning of relatively high-density patches in the population; and second, competition for light leads to a regular pattern of dominant, surviving trees.

The development of a regular pattern among surviving members of a plant population is highly suggestive of the action of intraspecific competition. As Ford pointed out³, however, it is difficult to prove a causal link between variations in plant growth and resource use when it is the assumed outcome of competition that is being studied. Two types of investigation can help to draw this link: simulation experiments with model systems; and field experiments in which some measure of resource use is obtained while spatial pattern is artificially manipulated.

Leps and Kindlmann described⁴ models of the development of spatial pattern over

time in even-aged plant populations. They used their simulation model to investigate the development of pattern over time in two situations: individual plant survival dependent on the competitive influence of neighbouring plants; and random mortality. They found that when survival depends on the influence of neighbours, spatial pattern becomes more regular over time whether the initial pattern is aggre-



Root-system sketches of *Haplopappus* (left) and *Chrysothamnus* (right) plants (drawn by C. Weber-Johnson). Late in the dry season, soil moisture changes most dramatically at 25 cm. This depth corresponds to a separation point for root biomass of the two species: *Haplopappus* lateral roots are concentrated in the upper, driest soil; *Chrysothamnus* laterals in the wetter soil. The thick taproots of *Chrysothamnus* can reach below 600 cm. (From ref. 5.)

gated or random. But when survival is independent of the influence of neighbours, spatial pattern approaches randomness from both aggregated and regular initial patterns. These results thus support the idea that intraspecific competition is the driving force for the development of a regular spatial pattern, but do not provide much insight on the nature of the competitive processes involved.

Desert shrub communities provide rather better opportunities for plant removal experiments than forest tree populations. Manning and Barbour have now investigated⁵ relationships between plant spatial pattern and resource use in two coexisting desert shrubs, *Haplopappus cooperi* and *Chrysothamnus teretifolius*. *Haplopappus* plants have a random pattern, *Chrysothamnus* plants an aggregated pattern, and individuals of the two populations are randomly distributed with respect to each other. For *Haplopappus*, removal of neighbouring plants of either or both species results in an increase in predawn xylem pressure potential (XPP), suggesting that competition for water has been relaxed. For *Chrysothamnus*, predawn XPP is not influenced by the removal of neighbours of either or

both species, suggesting that competition for water has not been affected by the experimental treatments. In this case, the above-ground spatial patterns reveal little about intra- or interspecific competition. The explanation for the competitive interactions within and between these two species seems to lie instead in their respective root morphologies (see figure).

In terms of Kenkel's model², the first phase of competition, for water, depends on the distribution of below-ground biomass. If the second phase, competition for light, is relatively unimportant, as seems likely in the desert-shrub community, and above-ground spatial pattern is determined by seed-dispersal mechanisms or the action of herbivores,

Manning and Barbour's results⁵ are not necessarily at odds with studies¹⁻⁴ associating intraspecific competition with the development of a regular pattern of surviving plants.

Plant populations experiencing intense intraspecific competition for resources exhibit a characteristic increase in mean biomass per plant (w) as the population density of surviving plants (d) declines. The 'self-thinning rule' is the empirical observation that in even-aged, single-species populations of genetically distinct plants, density-dependent mortality leads to the relationship $w = kd^{-3/2}$, where k is a constant. Although one of the earliest descriptions⁶ of the self-thinning rule provided an heuristic explanation based on distance to nearest neighbour as a measure of pattern, few attempts to account for the rule have embraced the spatial dynamics of self-thinning.

More recently, a re-evaluation⁷ of data on which the self-thinning rule is based suggests that self-thinning does not obey a single quantitative rule. Deviations from the proposed value of $-3/2$ for the exponent of the biomass-density relationship have been related⁸ to differences in the relative weights and dimensions of plant parts. It would be interesting to see how such variation is associated with changes in plant spatial pattern under self-thinning of the type identified by Kenkel. But, as Manning and Barbour's work reminds us, there is often more to plant spatial pattern than meets the eye. □

1. Veblen, T.T. in *The Ecology of Natural Disturbance and Patch Dynamics* (eds Pickett, S.T.A. & White, P.S.) 35-51 (Academic, Orlando, 1985).

2. Kenkel, N.C. *Ecology* **69**, 1017-1024 (1988).

3. Ford, E.D. *J. Ecol.* **63**, 311-333 (1975).

4. Leps, J. & Kindlmann, P. *Ecol. Mod.* **39**, 45-57 (1987).

5. Manning, S.J. & Barbour, M.G. *Am. J. Bot.* **75**, 885 (1988).

6. White, J. & Harper, J.L. *J. Ecol.* **58**, 467-485 (1970).

7. Weller, D.E. *Ecol. Monogr.* **57**, 23-43 (1987).

8. Weller, D.E. *Ecology* **68**, 813-821 (1987).

Gareth Hughes is at the Edinburgh School of Agriculture, University of Edinburgh, West Mains Road, Edinburgh EH9 3JG, UK.

Gene regulation

Homoeo boxes, POU proteins and the limits to promiscuity

Miranda Robertson

THE recent characterization of four proteins, three of them known mammalian transcriptional activators, may do much to diminish the lingering mystique of the homoeo domain and at the same time to focus attention on the role of protein-protein interactions in differential gene expression. The four proteins are Oct-2, which is produced in B lymphocytes where it binds to a conserved octamer sequence and activates immunoglobulin genes; Pit-1, which is synthesized in pituitary cells and implicated in the activation of growth-hormone and prolactin genes; unc-86, which has a genetically defined role in the differentiation of specific nematode neural cells; and Oct-1, a ubiquitous mammalian transcriptional activator of small nuclear RNA and histone

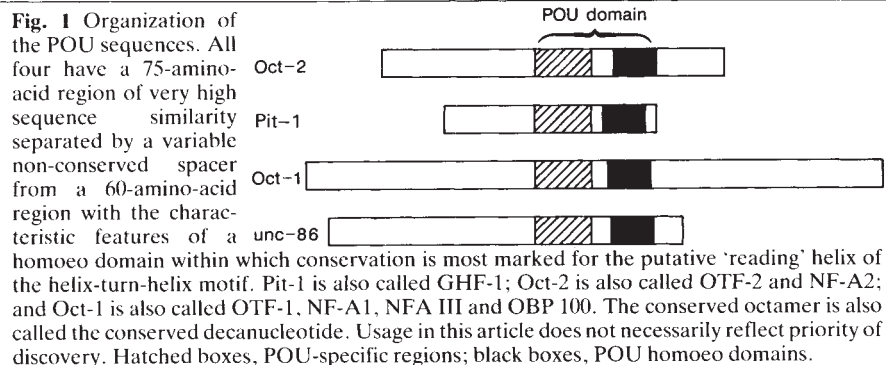
discoveries is to establish the homoeo motif definitively as a DNA-binding domain. The *Drosophila* genes in which the motif was originally discovered were defined genetically by their effects on morphogenesis, and although some of their products have since been shown to bind DNA elements implicated in gene regulation, their status as DNA-binding transcriptional regulators *in vivo* is still strictly speaking circumstantial. But Pit-1 and the two octamer-binding proteins were originally identified biochemically as transcriptional activators binding to identified DNA sequences, and it has been possible to show by deletion analysis of Oct-1 (ref. 6) and Oct-2 (refs 1, 5, 7) that the homoeo domain is essential for this specific binding.

POU proteins cannot be taken to reflect a common physiological function, what is their significance? By extending the deletion analysis of the POU domain to the conserved region outside the homoeo domain in Oct-1, Sturm and Herr (on page 601 of this issue¹²) show that it too contributes to DNA binding; a result that is consistent with the relaxed specificity shown by a truncated Pit-1 protein lacking part of the POU-specific region². Thus the similarities between the proteins should on the face of it indicate common or at least related DNA-binding specificities. This is largely borne out by the evidence so far; but it presents an interesting problem for the mechanism of gene activation, especially in the case of Oct-1 and Oct-2. These proteins have long been known to bind identical octamer sequences, yet only Oct-2 activates immunoglobulin genes. This implies that the important difference between the proteins lies in the areas outside the POU domain (where they are very different); and that in turn implies that the critical determinant of differential gene regulation is not the DNA-binding specificity of individual regulators but the interactions between them. What this means is that it should only very rarely be possible to impose the differentiated phenotype of one cell on another by the addition of one tissue-specific regulator, as the muscle-cell phenotype, for example, seems to be elicited by transfection with the *myoD* gene¹³. Further evidence for this comes from assaying the ability of Oct-1 and Oct-2 to activate different enhancer and promoter constructs in different cell types; and more indirectly from the character of the DNA-protein interactions of homoeo proteins, and the phenotype of unc-86 mutants.

POU phenotypes

Of the DNA-binding properties of unc-86 nothing is directly known because, like the *Drosophila* homoeo proteins, it was identified genetically. For the same reason, however, it is also the only one of the four POU proteins that is known to have pleiotropic effects: in two different neural lineages it specifies the differentiated phenotype of one of the two daughters of a dividing cell, so that in its absence the daughter cell retains the phenotype of the mother; and in the third case it is necessary for the maturation of a non-dividing neuron¹⁴. The differentiated phenotype is different in each case, so (assuming that, like the other POU proteins, unc-86 is a transcriptional regulator) Finney *et al.*⁸ conclude that its effects in each cell type differ by virtue of differences in the proteins with which it interacts.

Pit-1 should be illuminating about the mechanisms whereby closely related cells acquire distinct phenotypes, because it is expressed tissue-specifically in pituitary cells of two types: somatotrophs,



H2B genes that also acts through a conserved octamer motif.

Eight different research groups¹⁻⁸ have almost simultaneously reported the cloning and sequencing of the genes encoding these proteins (two of these reports appear on pages 544⁴ and 551⁵ of this issue), and four of them have got together to report⁹ on the remarkable sequence similarities between them. All four proteins have a 60-amino-acid homoeo domain with common sequence similarities that distinguish them from other known homoeo domains and define them as a new class; and all share a second region of sequence similarity spanning about 75 amino acids upstream of the homoeo domain (see Fig. 1, whose legend contains vital information on complexities of nomenclature not acknowledged in the text). Because these regions are shared by Pit-1, two octamer-binding proteins and unc-86, they have been designated a POU domain (pronounced "pow", the authors insist — understandably, given the alternative).

One immediate consequence of these

The second consequence of the sequence analysis is to put the significance of the homoeo motif in perspective. While many have pointed out from the beginning that homologies in molecular structure do not necessarily imply homologies of function at the level of the organism, much research on mammalian homoeo genes has nonetheless been driven by the supposition that, like the *Drosophila* genes, they may have a specific role in morphogenesis (see, for example, ref. 10).

Clearly, homoeo proteins need have no such role: none of the POU proteins is known to play a part in morphogenesis, and although Pit-1, Oct-2 and unc-86 all help determine differentiated phenotypes, Oct-1 performs a purely housekeeping function. (This is not to say however that homoeo proteins never contribute to mammalian morphogenesis: many of the *Drosophila* proteins have a morphogenetic role early in development and a tissue-specific one later¹¹, and the same could be true of mammalian proteins including the POU family.)

If the sequence similarities between the

which synthesize growth hormone; and lactotrophs, which synthesize prolactin^{2,3}, and is known to bind to the growth-hormone gene promoter. There is, however, at this stage some disagreement between Bodner *et al.*² and Ingraham *et al.*³ about its ability to bind to the prolactin promoter. Bodner *et al.* report that it does not, and the differentiation of lactotrophs therefore requires a second protein that activates the prolactin gene, as well as a mechanism for repressing the growth-hormone gene; Ingraham *et al.* report that it does bind the promoter, but agree with Bodner *et al.* on the need for a cell-type-specific repressor in lactotrophs; and they indirectly demonstrate this second point by showing that Pit-1 expressed from the cloned cDNA will not activate a co-transfectant growth-hormone promoter in a lactotroph cell line, though it will in HeLa cells. The resolution of the discrepancy will have to await further investigations by the two groups.

Homoeo protein promiscuity

If the similarity of the POU domains to one another indicates similar DNA-binding specificities, then it should follow that the conserved differences between these proteins and other homoeo proteins indicate different binding specificities. In fact, however, the differential binding specificities of different homoeo proteins do not seem to be very pronounced. Oct-2, for example, will bind to the yeast *Mata2* binding site¹ and one of the binding sites for the *Drosophila* segmentation gene *engrailed*^{1,5}, though with lower affinity than for its natural octamer site¹; and two other *Drosophila* proteins (Ubx and Abd') and Oct-2 will all activate a reporter gene in mammalian cells from either the natural *Drosophila* promoter sites or the mammalian octamer motif (page 598 of this issue¹⁵).

These experiments echo the results recently published by two groups^{16,17} reporting similar binding specificities between closely related *Drosophila* homoeo proteins which also seem capable of binding with lower affinity to the putative target sites of less closely related homoeo proteins. Much has been made^{18,19} of the possible contribution of competitive binding interactions between gene regulatory proteins to developmental decisions during embryogenesis, and Desplan *et al.*¹⁶ and Hoey and Levine¹⁷ point out that such interactions between closely related *Drosophila* proteins may account at least in part for early embryonic decisions in the fruitfly.

Competitive interactions can also take place between mammalian homoeo proteins; but the two most closely related mammalian homoeo domains yet identified are those of Oct-1 and Oct-2, whose identical binding specificities would seem to make competition inevitable, and yet

which do not apparently compete *in vivo*. If the differential actions of these proteins are not dictated by distinctive interactions with DNA then they must be defined by distinctive interactions with other proteins; and there are several indications that this is the case.

The limits to promiscuity

In the simplified picture of which Ptashne²⁰ has been the chief promulgator, transcriptional activators can be functionally divided into two surfaces: a DNA-binding surface that tethers the protein to a specific DNA sequence near a gene; and an activating surface that works by binding to a component of the transcriptional machinery and helping it to bind to the promoter. Activator regions are generally interchangeable for polymerase II promoters, and their strength depends upon intrinsic properties such as charge, and the distance of the protein from the promoter.

Within this framework, the simplest explanation for the distinction between Oct-1 and Oct-2 is that Oct-1 has a much weaker activating region and can therefore operate only over the very short distances between the octamer motif and the snRNA and H2B promoters, and not over the longer distances that separate the octamer from the promoter in immunoglobulin genes. Nothing is directly known

about the activating regions of the two proteins, though both have amino-terminal blocks of charged amino acids and glutamines characteristic of activating regions in other transcriptional activators; but LeBowitz *et al.*²¹ have shown that purified (not cloned) Oct-1 added in excess to HeLa cell nuclear extracts can activate an immunoglobulin heavy-chain promoter, which on the face of it supports the notion that Oct-1 is equivalent to Oct-2 in everything but strength.

Occam's razor is however not always the instrument of choice in biology²², and the simplest interpretation of this experiment is not the only one; nor is it the one the authors choose²¹, making the important point that Oct-1 in other contexts can also activate DNA replication and opting for an effect of "other cellular factors" to account for the multifarious activities of Oct-1 and the distinct actions *in vivo* of Oct-1 and Oct-2. Indeed as it turns out, their result is extremely difficult to reconcile with some more recent experiments²³ designed to test the role of protein-protein interactions in Oct-1 and Oct-2 activity, and which suggest that all activators are not interchangeable and the activating region of Oct-1 is inherently different from that of Oct-2. The experiments, which are depicted in Fig. 2, are based on the use of a synthetic enhancer consisting of multiple copies of the B element from the complex enhancer of simian virus 40 (SV40). The B element contains a conserved octamer that confers B-cell specificity on promoters linked to it; and the ability of Oct-1 and Oct-2 to activate transcription from the octamer was indirectly tested by introduction of two types of enhancer-promoter construct (Fig. 2a and b) into either HeLa cells (which contain only Oct-1) or lymphoid cells (which contain Oct-2).

The two promoters used by Tanaka *et al.*²³ were the β -globin promoter, which like those of immunoglobulin genes has the usual TATA box; and the snRNA promoter, which has a different type of promoter known as the proximal element (PE). The experiments showed that the SV40 octamer will activate the snRNA promoter in either HeLa cells or lymphoid cells, but the β -globin promoter only in lymphoid cells. This implies that the binding of Oct-1 can activate the snRNA promoter but not the β -globin promoter even when several copies are bound.

The results of these experiments could be explained if the activation domain of Oct-1 is not equivalent to that of Oct-2 but is specialized for promoters containing a PE, and as a further test of this possibility Tanaka *et al.* replaced the octamer upstream of the snRNA PE with binding sites for the yeast activator GAL4 (Fig. 2c). The GAL4 activator, which has a classical negatively charged activating domain, has been repeatedly shown to be

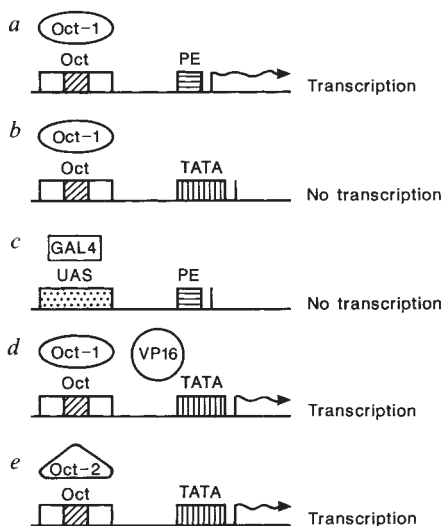


Fig. 2 Activation of small-nuclear (sn) RNA (PE) and β -globin (TATA) promoters through the SV40 octamer (a, b, d, e) or the GAL4 upstream element (UAS) (c) in HeLa cells (a-d) or lymphoid cells (e) in the presence of the herpesvirus transcriptional activator VP16 (d) or GAL4 (c). The conserved octamer sequence in the SV40 element in fact straddles two *sph* motifs through which the β -globin promoter can be activated in HeLa cells. These were inactivated in the experiments shown by point mutations that left the octamer intact. Hatched areas, SV40 enhancer element octamer sequence; stippled areas, upstream activating sequence (UAS) required to bind GAL4; horizontal stripes, snRNA proximal element (PE) promoter; vertical stripes, β -globin TATA-box promoter.

capable of activating mammalian genes of diverse kinds; but failed to activate the snRNA promoter, suggesting that indeed promoters containing PEs require an activator of a different kind.

This however presents a paradox, because Oct-1 can also activate the H2B promoter which has a TATA box. To explain this, Tanaka *et al.* postulate an interaction between Oct-1 and a second protein that provides the necessary TATA activating region; from which they reason that it should be possible to enable Oct-1 to activate the β -globin TATA by supplying it with an appropriate activating domain.

The means for achieving this conversion lay to hand in the viral protein VP16, a powerful activator which does not bind DNA itself²⁴ but acts via cellular DNA-binding proteins that have recently been shown to include Oct-1²⁵. In the presence of VP16, the β -globin promoter could be activated through the SV40 octamer in HeLa cells. If, however, as this implies, Oct-1 has no TATA activator of its own, the LeBowitz *et al.* experiment should not have worked, and can be explained only by some adventitious interaction made possible by the particular conditions *in vitro*; and in any case parsimony seems a lost cause.

Further evidence for the critical importance of differential protein interactions in determining the behaviour of transcriptional activators comes from investigations by Müller *et al.*⁴ on the actions of Oct-2 bound to the two conserved octamers in the control region of the immunoglobulin heavy-chain gene. One of these octamers is in the promoter, about 20 base-pairs from the TATA box, the other is in the enhancer, some 2,000 base pairs distant. Müller *et al.* have been able to show that cloned Oct-2 is sufficient to activate a co-transfectant reporter gene from the promoter octamer, which is consistent with earlier evidence that Oct-2 is responsible for cell-type specific transcription from that promoter^{26,27}; but surprisingly, it will not activate from the immunoglobulin enhancer element in similar experiments.

This may be because the two octamers are bound *in vivo* by two different Oct-2 proteins generated by alternative splicing from the Oct-2 gene: one, Oct-2A, seems to operate from the immunoglobulin promoter⁴, while the other, Oct-2B, may mediate activation from the enhancer²⁸. It is unlikely that the difference between the two proteins lies simply in the strength of their activating regions — Oct-2A being too weak to act from the enhancer site — because Oct-2A is inactive even from a synthetic multimeric enhancer (Patrick Matthias and Walter Schaffner, personal communication) whose multiplicity might be expected to overcome the weakness of the individual activating regions. The distinction thus seems likely to lie in different

protein-protein interactions that may be a function partly of the differences between the Oct-2A and Oct-2B proteins, and partly of the different context of the octamer in the promoter and the enhancer, which may contribute to the cooperative binding of other regulatory proteins.

Further clues and conundra may lurk in the non-POU regions of the Oct-2 protein sequence. There is for example a leucine-rich region with the heptad repeat characteristic of the (still) putative leucine zipper implicated in the formation of DNA-binding dimers²⁹ and which suggests, therefore, the possibility of interactions that foster cooperative binding with other proteins that might modify the activity of the Oct-2 monomer.

More curiously, Clerc *et al.* draw attention to a 278-amino-acid open reading frame overlapping with the coding region for the carboxy terminus of Oct-2. They cryptically remark that the sequence has features in common with the adenoviral activator protein E1A, which like VP16 seems to be a freelance activating domain. Is it conceivable that as B lymphocytes enter terminal differentiation as plasma cells committed to spewing thousands of immunoglobulin molecules into the bloodstream, they switch on the synthesis of a 278-amino-acid activator that empowers Oct-2 to monopolize the cell's transcriptional machinery for the task? □

1. Ko, H.-S., East, P., McBride, W. & Staudt, L.M. *Cell* **55**, 135-144 (1988).
2. Bodner, M. *et al.* *Cell* **55**, 505-518 (1988).
3. Ingraham, H.A. *et al.* *Cell* **55**, 519-529 (1988).
4. Müller, M.M., Ruppert, S., Schaffner, W. & Matthias, P. *Nature* **336**, 544-551 (1988).
5. Scheidter, C. *et al.* *Nature* **336**, 551-557 (1988).
6. Sturm, R.A., Das, G. & Herr, W. *Genes Dev.* **2**, 1582-1599 (1988).
7. Clerc, R.G., Corcoran, L., LeBowitz, J.H., Baltimore, D. & Sharp, P.A. *Genes Dev.* **2**, 1570-1581 (1988).
8. Finney, M., Ruvkun, G. & Horvitz, R. *Cell* (in the press).
9. Herr, W. *et al.* *Genes Dev.* **2**, 1513-1516 (1988).
10. Giehring, W.J. *Science* **236**, 1245-1252 (1987).
11. Ingham, P. *Nature* **335**, 25-34 (1988).
12. Sturm, R.A. & Herr, W. *Nature* **336**, 601-604 (1988).
13. Tappin, S.J. *et al.* *Science* **242**, 405-411 (1988).
14. Desai, C., Garriga, G., McIntyre, S.L. & Horvitz, R.H. *Nature* (in the press).
15. Thali, M., Müller, M., DeLorenzi, M., Matthias, P. & Bienz, M. *Nature* **336**, 598-601 (1988).
16. Desplan, C., Theis, J. & O'Farrell, P.H. *Cell* **54**, 1081-1090 (1988).
17. Hoey, T. & Levine, M. *Nature* **332**, 858-861 (1988).
18. Ptashne, M. *A Genetic Switch* (Cell/Blackwell Scientific, Palo Alto, 1986).
19. Robertson, M. *Nature* **327**, 556-557 (1988).
20. Ptashne, M. *Nature* **335**, 683-689 (1988).
21. LeBowitz, J.H., Kobayashi, T., Staudt, L., Baltimore D. & Sharp, P.A. *Genes Dev.* **2**, 1227-1237 (1988).
22. Crick, F.H.C. *What Mad Pursuit* (Basic Books, New York, 1988).
23. Tanaka, M., Grossniklaus, U., Herr, W. & Hernandez, N. *Genes Dev.* **2**, 1764-1778 (1988).
24. Triceberg, S.J., La Marco, K.L. & McKnight, S.L. *Genes Dev.* **2**, 730-742 (1988).
25. Gerstner, T. & Roeder, R.G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6347-6351 (1988).
26. Wirth, T., Staudt, L. & Baltimore, D. *Nature* **329**, 174-177 (1987).
27. Dreyfus, M., Doyen, N. & Rougeon, F. *EMBO J.* **6**, 1685-1690 (1987).
28. Schreiber, E., Matthias, P., Müller, M.M. & Schaffner, W. *EMBO J.* (in the press).
29. Landshulz, W.H., Johnson, P.F. & McKnight, S.L. *Science* **240**, 1759-1764 (1988).

Miranda Robertson is Biology Editor of Nature.

Daedalus

Let the money talk

THE capitalist mode of technical development is very persuasive. In an uncertain world, the shrewdest technical forecasts are probably made by informed investors risking their own money. So the promoters of a new scheme issue a prospectus appealing for funds, and the market decides.

But these days, says Daedalus, almost every technological project has drawbacks. For every British exporter longing for the Channel Tunnel, for example, there may be some landowner facing the ruin of his land by its approach roads. Capitalism is therefore imperfect. The exporter's money can talk in the market: he can buy shares in the Tunnel. But the landowner's money cannot talk for him. He must appeal to environmental-protection or social-audit agencies, or stir up political protests.

In this connection, Daedalus recalls that brilliant social invention, the protection racket. If our landowner had the chance to buy protection against the Channel Tunnel, his real opposition to it would soon become apparent: and in sound financial terms, too. So Daedalus proposes that a company should be able to issue not merely positive shares to promote a project, but negative ones purchased as 'protection' against it.

This appealing scheme has an immediate snag. A classic protection racket is self-perpetuating. Once you have paid up for protection, the same (or even another) gang can come back and ask to be paid again. To prevent this unfair trick, Daedalus's negative shares are formally an investment, not merely in abandoning the project, but actually making it illegal. So, sadly but inevitably, the Government must be brought in. Daedalus envisages a 'Government Obstruction Agency' (a natural home for ardent bureaucrats) with the job of overseeing the issuing of negative shares.

This technological decision-making will acquire a pleasing symmetry. Any new promoter will probably issue both positive and negative shares, and judge from their uptake whether to set up the project in earnest, or run it as a protection racket and hope to be bought off. On the one hand is costly and uncertain development, with any ultimate profit delayed maybe for years; on the other is a swift pay-off in protection money, smaller but immediate. The investing public, voting with its wallet, will decide. Costs and benefits will both be judged by the public in sound financial terms. The major flaw in modern capitalism, its insensitivity to social and environmental drawbacks, will be healed. And when at last every new technical idea turns out to be more profitable as a protection racket than a project, why, the world will have enough technology, and progress will automatically stop. David Jones

Origin of mutants disputed

SIR—Cairns *et al.*¹ present experiments that lead them to the provocative conclusion that, in response to a particular environmental challenge, bacteria generate a class of mutations that specifically enhance their fitness in the novel environment. This conclusion is at variance with the neo-darwinian orthodoxy that mutations occur strictly at random with respect to any such challenge². We here refer to these alternative possibilities as the 'directed mutation' model and the 'spontaneous mutation' model, respectively. Support for the directed mutation model was derived from the observation that mutation rates to variants with enhanced growth on a novel medium are apparently higher under exposure to the medium than without exposure. We suggest that the results as presented are consistent with both models, so that additional evidence is desirable before the spontaneous mutation model, so strongly supported by previous experiments^{1,3}, can be rejected.

The data presented by Cairns *et al.* are of two kinds. First, they use the statistical procedure of the fluctuation test³⁻⁵ to estimate the distribution among replicate cultures of the numbers of mutations conferring the ability to ferment lactose, in an initially Lac⁻ strain of *Escherichia coli*. Second, they monitor the time of appearance of these Lac⁺ mutations after accumulation in the absence of lactose, followed by its addition. The mutations detected in these experiments are of two kinds, suppressors of a chain-terminating (amber) mutation in the *LacZ* structural gene, and mutations to the Lac⁺ phenotype which are due to loss of an inserted sequence upstream of *LacZ*.

In their interpretation of the data, the authors assume that all Lac⁺ mutations were neutral in the absence of lactose. This assumption implies that, on the spontaneous mutation model, the observed distribution of mutations should be the same whether or not the bacteria were previously exposed to medium containing lactose. If some of the Lac⁺ mutations were deleterious on medium lacking lactose, however, those mutations would have failed to propagate themselves as fast

as neutral Lac⁺ mutations. In that case, the number of Lac⁺ mutations would be higher in bacteria exposed to lactose-rich medium because that medium would enable a class of spontaneous mutations to persist that would otherwise be lost. The data presented by Cairns *et al.* do not seem to permit us to reject this possibility, which is fully in line with the spontaneous mutation theory. One test of this idea would be to generalize the distributions of numbers of mutations derived by Lea and Coulson³ for the fluctuation test to include selection on the mutants. This has been done by Koch⁵, and Lenski, Slatkin and Ayala have shown (personal communication) that it can give a quantitative fit to the curves found by Cairns *et al.* In the second series of experiments, the delayed appearance of the Lac⁺ mutants in the absence of lactose is evidently consistent with their being at a selective disadvantage unless lactose is present.

In considering this interpretation, it should be borne in mind that because of the nature of the Lac⁻ mutations that were used, the majority of the Lac⁺ mutants in the experiments of Cairns *et al.* are probably not simple reversions to wild-type, that is they do not restore the ancestral Lac⁺ sequence of the *LacZ* region, but involve amber suppression or rearrangement of upstream regions. There is much evidence that the majority of spontaneous mutations with observable phenotypic effects are deleterious^{2,6}, and this is known to be the case for chromosome rearrangements without major loss of genetic material⁷; from the way in which suppression of chain-terminating mutations works, it is extremely likely that these would also have deleterious effects on fitness^{8,9} and indeed there is some direct evidence for this¹⁰⁻¹². This view of these experiments may also explain the difference noted by Cairns *et al.* between their results using nutritional markers and those previously obtained by others using phage and drug resistance markers^{1,3}. Such mutations are known to show a delay of several cell generations in expression, after their initial occurrence¹, and it seems reasonable that this might apply to any deleterious effects that they may also have. If this were true, these mutations would indeed behave as truly neutral, at least until the time when their phenotypic effects begin to be expressed.

What evidence might be used to discriminate between the two models? If it is found that the mutations appearing only after exposure to the lactose-rich medium are selectively inferior to the ancestral Lac⁺ bacteria when tested on medium lacking lactose, the spontaneous mutation model would be supported. Highly sensitive techniques are available for

measuring the relative fitnesses of different bacterial genotypes^{6,7,12} and these could be applied to the Lac⁺ mutations in question. There even seems to be some internal evidence in Fig. 3 of Cairns *et al.* for a selective disadvantage to these mutations. In this figure, which shows the results of their second series of experiments, the plates that were overlaid with lactose-containing medium earliest can be seen to have the highest numbers of Lac⁺ cells at all later times, which suggests that the longer plates were kept without lactose, the lower the number of surviving Lac⁺ cells. Even if the late-appearing mutants do not exhibit such inferiority, the spontaneous mutation model can safely be rejected only if it can be demonstrated that natural selection among descendants of each late-appearing mutant could not have improved their fitness; such improvement would be likely to occur, given sufficient time¹².

D. CHARLESWORTH
B. CHARLESWORTH

Department of Ecology and
Evolution,
University of Chicago,
915 East 57th Street, Chicago,
Illinois 60637, USA

J.J. BULL

Department of Zoology,
University of Texas, Austin,
Texas 78712, USA

SIR—Loci at which directed mutation takes place (see Cairns *et al.*¹) are likely to be 'heritable soma'. That is, although the contents are inherited, the information they contain is not potentially immortal, and so the loci are not germ plasm. Mutations at those loci will therefore have the same logical role in evolution as somatic adaptation in multicellular organisms, such as the strengthening of our muscles when we exercise.

Consider an example of directed mutation where the mechanism is well known. In the fission yeast, *Schizosaccharomyces pombe*, the contents of the locus *mat1* determine whether the mating-type of the cell is plus or minus². The contents of this locus are regularly replaced by a copy of the contents of either *mat2* or *mat3*, loci which contain silent master copies of the alleles determining plus or minus mating-type respectively. The locus *mat1* is not germ plasm, because a point mutation at *mat1* will have a short-lived effect before the whole allele is replaced and subsequently degraded.

DNA is not only the repository of inherited and potentially immortal information — it is also just another biochemical tool that can be used in the service of the cell. No doubt cells switch states using a variety of mechanisms: protein concentrations, methylation of DNA, condensation of DNA, and sometimes the base sequence of DNA. No great theoretical

1. Cairns, J., Overbaugh, J. & Miller, S. *Nature* **335**, 142-145 (1988).
2. Muller, H.J. in *Genetics, Paleontology and Evolution* (eds Jepsen, G.L., Simpson, G.G., & Mayr, E.) 421-445 (Princeton University Press, Princeton, 1949).
3. Luria, S.E. & Delbrück, M. *Genetics* **28**, 491-511 (1943).
4. Lea, D.E. & Coulson, C.A. *J. Genet.* **49**, 264-285 (1949).
5. Koch, A.L. *Mutat. Res.* **95**, 129-143 (1982).
6. Lenski, R.E. *Evolution* **42**, 425-432 (1988).
7. Hill, C.W. & Gray, J.A. *Genetics* **119**, 771-778 (1988).
8. Garen, A. *Science* **160**, 149-150 (1968).
9. Ozeki, H. *et al.* in *Transfer RNA: Biological Aspects* (eds Söll, D., Abelson, J.N. & Schimmel, P.R.) 341-362 (Cold Spring Harbor Press, Cold Spring Harbor, 1980).
10. Soll, L. *Proc. natn. Acad. Sci. U.S.A.* **63**, 392-399 (1969).
11. Hill, C.W., Foulds, J., Soll, L. & Berg, P. *J. molec. Biol.* **39**, 563-581 (1969).
12. Lenski, R.E. *Evolution* **42**, 433-440 (1988).

consequences can flow from this practical choice of means.

Mechanisms of directed variation in general can be expected to be similar to the yeast example. To evolve, such a mechanism must have been used many times in the ancestral lineage of any modern cell, and must still be used frequently enough for selection to be effective against mutation at the loci underlying the mechanism. A directed locus must therefore have been mutated many times, and must still be regularly mutated. A locus subjected to so much change is likely to qualify as heritable soma rather than germ plasm, as any mutation is likely to be 'overwritten' fairly rapidly by directed mutation.

E. coli may be professional mutators, routinely modifying their DNA to switch between metabolic states. The immune systems of vertebrates contain cells that create antibodies using high rates of somatic mutation³, thus effectively segregating their 'scratchpad DNA' from their germ plasm DNA in a way unavailable to unicellular organisms. It seems likely that whatever mechanism underlies the *E. coli* results of Cairns *et al.*, the cases will be logically parallel, and have the same evolutionary status.

I have suggested elsewhere⁴ that the centrosome may contain inherited and potentially immortal information that is not encoded by nucleic acids. If this were true, then the conventional identification of DNA with germ plasm would have to be modified on the dual grounds that not quite all chromosomal DNA is germ plasm, and not quite all germ plasm is DNA.

ALAN GRAFEN

*Animal Behaviour Research Group,
Zoology Department, Oxford University,
Oxford OX1 3PS, UK*

1. Cairns, J., Overbaugh, J. & Miller, S. *Nature* **335**, 142–145 (1988).
2. Klar, A.J.S. *Nature* **326**, 466–470 (1987).
3. Alberts, B. *et al.* *Molecular Biology of the Cell* (Garland, New York, 1983).
4. Grafen, A. *J. theor. Biol.* **131**, 163–173 (1988).

SIR—In their article, Cairns *et al.*¹ present results and an interpretation which have already aroused considerable interest and controversy. It is a guiding principle in science that a radical new interpretation (in this case, one invoking the inheritance of acquired characteristics) should only be considered if simpler explanations based on existing knowledge are inadequate. In the same issue, Stahl² invites readers to devise their own explanation, and puts forward one of his own. We present here an even simpler possibility for some, but not necessarily all, of the observations of Cairns *et al.*, which is in keeping with current knowledge of the genetics and molecular biology of *E. coli*.

Whereas replication of DNA is known to be highly accurate (in the range of 10^{-9} – 10^{-10} errors per base per generation), the

accuracy of transcription and translation is far less. Several studies have shown that the error levels are of the order of 10^{-4} for RNA bases during transcription¹ and as high as 10^{-3} random errors per amino-acid residue during translation⁵. From studies on the *lac*^{amber} mutations in the gene for β -galactosidase (one of the genes studied by Cairns *et al.*) it is clear that the mutants still contain measurable residual enzyme activity^{1,6} and that some growth in the presence of lactose is to be expected. It should be noted that in this situation the cells which are most error-prone will grow fastest and that any introduction of extra random noise in the pathways of information transfer will increase the probability of mutations from *lac*[–] to *lac*⁺ (ref. 7).

The *lac*^{amber} cells attempting to grow in the presence of lactose will be severely limited for energy and biosynthetic substrates. We would argue that such limitations could introduce additional random noise into the information transfer process and, together with the basal error levels, may be responsible for at least some of the effects that Cairns *et al.* discuss. The accuracy of DNA³, RNA⁸ and protein⁵ synthesis at any particular site depends on the pool sizes of the correct and related but incorrect substrates and on the availability of energy for repairing wrong insertions. If the relative concentrations of the correct substrate for a particular site in protein or nucleic-acid synthesis are reduced, the error frequency rises. As errors increase a feedback mechanism may be set up and, as described in the general error theory⁷, give rise to an increased number of mutations.

As we mentioned, this proposal may not explain all the results reported by Cairns *et al.* but it will account for a significant deviation from the Luria and Delbruck distribution of mutants in a population. To decide how much an error feedback contributes to the appearance of late mutants, one needs much more and more detailed data on the non-randomness of the mutations. Although Cairns *et al.* make non-randomness the central theme of their arguments, it seems fair to say that the data on this is perhaps the least satisfactory part of their presentation. It would seem important to know the relative frequencies of *lac*⁺ and *val*^R mutants in their strain. In particular the reversion frequency of *lac*^{amber} can vary greatly with the site of the nonsense codon, which they do not specify. If the nonsense codon is in a part of the enzyme where the nature of the amino acid inserted in its place does not affect activity, mutations in many transferRNAs will give *lac*⁺ mutants. Should this lead to a markedly higher frequency of *lac*⁺ than *val*^R mutants, one may need to detail the mutant numbers and statistical analysis (not given) to demonstrate non-randomness.

In general, under conditions where the

disabled strain will grow but only very poorly, there is first likely to be weak phenotypic suppression and later mutation, but this will not happen when the mutation would not be selected. If this is correct, it will account for observed deviations from Luria and Delbruck expectations. The involvement of errors could be tested by manipulating translational or transcriptional accuracy.

R. HOLLIDAY

*CSIRO Laboratory for Molecular Biology,
North Ryde, Sydney, Australia*

R.F. ROSENBERGER

*Genetics Division,
National Institute for Medical Research,
Mill Hill, London NW7 1AA, UK*

1. Cairns, J., Overbaugh, J. & Miller, S. *Nature* **335**, 142–145 (1988).
2. Stahl, F.W. *Nature* **335**, 112–113 (1988).
3. Fersht, A.R., Knill-Jones, J.W. & Tsui, W.C. *J. molec. Biol.* **156**, 37–51 (1982).
4. Rosenberger, R.F. & Hilton, J. *Molec. gen. Genet.* **191**, 207–212 (1983).
5. Kurland, C.G. & Ehrenberg, M. *A. Rev. Biophys.* **16**, 291–317 (1987).
6. Gallant, J., Ehrlich, H., Weiss, R., Palmer, L. & Nyari, L. *Molec. Gen. Genet.* **186**, 221–227 (1982).
7. Kirkwood, T.B.L., Holliday, R. & Rosenberger, R.F. *Int. Rev. Cytol.* **92**, 93–132 (1984).
8. Blank, A., Gallant, J.A., Burgess, R.R. & Loeb, L.A. *Biochemistry* **25**, 5920–5928 (1986).

SIR—The interesting paper by Cairns *et al.*¹ provided presumptive evidence for bacterial mutations occurring adaptively rather than randomly with respect to selection. However, their main evidence, on mutation to lactose fermentation, may have a more conventional interpretation.

The authors assume that there is no natural selection in their cultures. If one relaxes this implausible assumption the expectation of a Poisson distribution of mutant frequencies no longer holds; the direction of change in the expectation is the same as that found.

The most likely source of most of the selection against *Lac*⁺ revertants is periodic selection². This refers to positive selection for fitter variants arising in a population. These variants are more likely to arise in bacteria with the majority genotype, with the result that rarer alleles will be eliminated as the fitter variants take over the population.

The data presented by Cairns *et al.* can be explained if *Lac*⁺ mutations accumulate at the rate seen in lactose agar for about half a day between the latest wave of periodic selection for a fitter variant and the time of plating on lactose agar. Whether this actually happened or not I do not know, but in this case it still needs to be shown that the apparent unicorn is not a goat after all.

LEIGH M. VAN VALEN

*Department of Ecology and Evolution,
University of Chicago,
915 East 57 Street, Chicago,
Illinois 60637, USA*

1. Cairns, J., Overbaugh, J. & Miller, S. *Nature* **335**, 142–145 (1988).
2. Atwood, K.C., Schneider, L.H. & Ryan, F.J. *Cold Spring Harb. Symp. quant. Biol.* **16**, 345–355 (1951).

SIR—Cairns *et al.* (*Nature* **335**, 142–145; 1988) have recently argued that specific mutations may be induced in response to selection. Their case was based in part on the excess of mutations arising in stationary phase and under selective conditions relative to those arising during exponential growth in the absence of selection.

These experiments may not provide definitive evidence for directed mutation, as it is well known that the mutation rate of *E. coli* varies with environmental conditions, and in a way that may vary depending on the DNA composition in the vicinity of the mutation. This makes the control experiment very important and it is unfortunate that the control used by Cairns *et al.* was a gene for valine resistance. Many of these are frameshift mutations and may therefore not be directly comparable to the mutations that revert a nonsense mutation used by Cairns *et al.* to test for mutations induced in response to selection.

A. DANCHIN

*Institut Pasteur,
28 Rue du Dr Roux,
75724 Paris, Cedex 15, France*

SIR—I would like to suggest an alternative explanation for the observations that led Cairns *et al.* (*Nature* **335**, 142–145; 1988) to suggest there may be a mechanism for a cell to mutate specifically in response to selection pressure. If the Lac⁺ mutant bacteria have a slower growth rate than their Lac⁻ parents, then the contribution of mutations occurring early in the growth of a culture will be curtailed. This will have the effect of producing a distribution similar to that observed by Cairns *et al.*, as I have shown in simulations using an explicit model.

My colleagues S.-K. Liu and I.-S. Hwang (personal communication) have examined the distribution of the number of Gal⁺ revertants of Gal⁻ parents in *E. coli* and show a deviation from the prediction of Luria–Delbruck clonal theory similar to that seen by Cairns *et al.* In this case more than half the Gal⁺ mutants do grow relatively slowly, which could account in part for the deviation from the expected distribution.

IRWIN TESSMAN

*Department of Biological Sciences,
Purdue University,
West Lafayette,
Indiana 47907, USA*

CAIRNS REPLIES—Before responding to certain specific comments, it might be helpful to remind readers of the context of our work. Because most mutations are rare, the circumstances surrounding their occurrence can be studied only in very large populations, such as bacterial cultures. Some very elegant experiments performed in the 1940s and 1950s showed that certain kinds of mutation in bacteria arise spontaneously, before there has been any selection. That conclusion

agreed so well with everyone's preconceptions that the question of the origin of mutants has been left virtually untouched since then. We resurrected the issue partly because of increasing discomfort with the conventional view of the origin of the genetic changes in cancer¹ and in part because we realized that the classical experiments had not been a fair test. In our paper², we describe a study of three types of mutation *E. coli*.

The first is the reversion of a nonsense mutation in *lacZ*. We discussed this case at some length, even though the results were not at all clear-cut, because it shows how the timing of mutational events can be deduced from the distribution of numbers and time of appearance of mutant colonies. We found that although many of the mutations occur during the prior growth of each culture (before there has been any selection), most mutants seem to arise later, after the bacteria were put on to selective lactose-minimal plates. As a result, the distribution of mutant numbers among different cultures is much narrower than expected.

Our critics have offered several explanations for this result. Van Valen suggests that mutants are being lost all the time through what is called "periodic selection" (namely, the loss of the accumulated mutants in a population due to takeover by a fitter variant, which is observed in an irregular fashion when cultures are maintained for long periods under sub-optimal conditions); since we are observing a regular phenomenon in cultures following rapid growth for a relatively short period of time, this explanation is most implausible.

Several people have suggested that the Lac⁺ mutants are growing more slowly or surviving less well than non-mutants (ref. 3; the letters from Charlesworth *et al.* and from Tessman; R.E. Lenski, M. Slatkin and F.J. Ayala, personal communication). We feel that the simple version of this hypothesis is not very likely because most of the mutants we detected behaved as if they were revertants to wild type and, by the time that we could test their growth rate, appeared to grow as well as non-mutants. But there is lurking in this suggestion the more complex idea that it may be a special property of all new mutants that they are temporarily at a great disadvantage in the absence of positive selection pressure, and that, of course, is just another way of describing the very anomalies we are seeking to explain.

Our experiments with *lacZ* had a second part, in which we looked at the late accumulation of Lac⁺ revertants on lactose-minimal plates. We showed that extra revertants do not accumulate unless lactose is present, and that this effect is specific because it does not cause the accumulation of unselected mutations (to valine-resistance). From this we con-

cluded that the *lacZ* mutation is not significantly 'leaky', our *lacZ* strain is not growing on lactose-minimal plates and lactose is not acting as an indiscriminate mutagen.

It is imaginable, however, that the frameshift mutations, which are commonly the cause of valine-resistance, may not respond to starvation in the same way as point mutations, as Danchin suggests above. This is a valid criticism, and perhaps we should have used as control the reversion of point mutations in other genes more like *lacZ*. (Note, however, that the issue is whether our *lacZ* mutation is leaky rather than what might possibly be the reasons for its leakiness, as Holliday and Rosenberger seem to think.) The experiment has also been criticized on the grounds that the steady decline in the number of Lac⁺ and valine-resistant mutants in the absence of selection is additional evidence that all mutants are at a disadvantage (see ref. 3; the letter from Charlesworth *et al.*; R.E. Lenski, M. Slatkin and F.J. Ayala, personal communication). But an unpublished control, using deliberately introduced Lac⁺ cells, showed that this decline is not due to death of cells in the absence of an energy source, but to a slight delay in colony formation by cells that have been deprived of energy for several days.

The *lacZ* experiments led us to the rather cautious conclusion that populations of bacteria in stationary phase seem to have "some way of producing (or selectively retaining) only the most appropriate mutations". But we too were worried about the background noise, due to the sizeable contribution from spontaneous mutation during prior growth of the cultures. Indeed, if these had been the only experiments, the paper would not have been written.

We did, however, discuss at length two examples of genetic events that apparently occur only under conditions of selection, and it was in connection with these examples that we suggested that bacteria may be able to determine which mutations occur. Since our paper was written, other such examples have come to light^{4,5}. Perhaps these will stimulate a second round of correspondence, specifically directed to mutations that occur only when there is selection.

In the meantime, we should remember that the doctrine, so vehemently defended by our critics, is an essentially negative assertion: phenotype never comes before genotype, and so all mutations have to arise before there can be any intimation of their consequences. Yet, it is easy to imagine mechanisms that might test the utility of mutations before they become irrevocably fixed into the genome⁶. If this seems too heretical, the heresy can be softened by postulating that the licence to indulge in such games of trial-and-error

might be confined to limited regions of the genome, though Grafen thinks that our heresy did not go far enough, because we seemed to assume that any form of directed mutation has to start with random changes, and so we were still neo-darwinists at heart.

This letter is primarily a response to our critics. Several commentators, however, have written to say that our experiments have tended to conform their own beliefs and interested readers should turn to their papers⁷⁻¹⁰ for the details.

We should end with one last thought. We now know that, in the processing of biological information, almost anything is possible. Sequences are spliced, rearranged, cast aside, resurrected, and to a limited extent may even be invented when the need arises, and so it should not be difficult for an organism to devise a way of testing phenotype before adopting any new genotype. It therefore seems almost perverse to maintain, as a matter of principle, that such a mechanism has never evolved.

JOHN CAIRNS

Harvard School of Public Health,
Boston, Massachusetts 02115, USA

1. Cairns, J. *Nature* **289**, 353-357 (1981).
2. Cairns, J., Overbaugh, J. & Miller, S. *Nature* **335**, 142-145 (1988).
3. Partridge, L. & Morgan, M.J. *Nature* **336**, 22 (1988).
4. Hall, B.G. *Genetics* (in the press).
5. Benson, S.A. *Nature* **336**, 21-2 (1988).
6. Stahl, F.W. *Nature* **35**, 112-113 (1988).
7. Wright, B.E. *C.r. Trav. Lab. Carlsberg Ser. Physiol.* **25**, 173-211 (1953).
8. Fitch, W.M. *Evolution* **36**, 1133-1143 (1982).
9. Shapiro, J.A. *Mol. gen. Genet.* **194**, 79-90 (1988).
10. Opadia-Kadima, G.Z. *J. theor. Biol.* **124**, 127-135 (1987).

Calcium channels

SIR—In a recent News and Views article, Rink discussed¹ 'real' receptor-operated calcium channels (ROCCs) on the basis of data demonstrating single Ba²⁺-selective channels obtained by incorporating membrane vesicles from thrombin-stimulated platelets into planar bilayers². It is not clear from the experiments described² whether the channels studied are derived from the plasma membrane, and no direct link has been established between agonist-receptor interaction and channel opening. It may therefore be premature to claim that an ROCC has been found.

It would be helpful if a distinction could be maintained not only between voltage-sensitive and insensitive Ca²⁺ channels, but also between activation of cellular Ca²⁺ entry mediated by second messengers (such as inositol-1,4,5-trisphosphate (Ins(1,4,5)P₃)/inositol-1,3,4,5-tetrakisphosphate (Ins(1,3,4,5)P₄), refs 3,4) and direct or G-protein-linked Ca²⁺ channel opening^{5,6}. The former could be called second-messenger-operated calcium channels (SMOCCs) and the term ROCC should be reserved for the latter category.

It is possible to have a 'proper' voltage-gated Ca²⁺ channel operating as an

SMOCC. In insulin-secreting cells, carbohydrate stimulation depolarizes the cell membrane and opens voltage-gated Ca²⁺ channels of the 'L' type, but the small depolarization cannot by itself activate these channels without additional second messenger-mediated modulation, probably resulting from metabolic generation of diacylglycerol⁷.

Rink claims¹, referring to the recently published work of Penner *et al.* (ref. 8), that the role of Ins(1,3,4,5)P₃ in the control of Ca²⁺ entry remains controversial. But the interesting experiments of Penner *et al.* were not designed for a critical test of the hypothesis that a combined action of Ins(1,4,5)P₃ and Ins(1,3,4,5)P₄ is needed for Ca²⁺ uptake^{4,9}, and the authors point out that their results cannot exclude a role for Ins(1,3,4,5)P₃. A test of the Ins(1,4,5)P₃/Ins(1,3,4,5)P₄ hypothesis requires a preparation such as the one we have used⁴, in which agonists consistently evoke Ca²⁺ influx and in which the same single cells can be internally perfused with both control and test solutions.

There is now clear evidence for voltage-sensitive as well as voltage-insensitive Ca²⁺ channels. Both types can be controlled by receptor-activation directly, via G-proteins, or indirectly, via second messengers such as inositol polyphosphates or diacylglycerol. The demonstration that one of many possible Ca²⁺-gating mechanisms is absent in one cell type cannot, of course, be taken as evidence against a particular mode of operation found in another system.

O. H. PETERSEN

The Physiological Laboratory,
University of Liverpool, PO Box 147,
Liverpool L69 3BX, UK

1. Rink, T.J. *Nature* **334**, 649-650 (1988).
2. Zschauer, A., Van Breemen, C., Buhler, F.R. & Nelson, M.T. *Nature* **334**, 703-705 (1988).
3. Berridge, M.J. *ISI Atlas of Science: Pharmacology* **1**, 91 (1987).
4. Morris, A.P., Gallacher, D.V., Irvine, R.F. & Petersen, O.H. *Nature* **330**, 653-655 (1987).
5. Benham, C.D. & Tsien, R.W. *Nature* **328**, 275-278 (1987).
6. Hescheler, J. *et al.* *EMBO J.* **7**, 619-624 (1988).
7. Petersen, O.H. *ISI Atlas of Science: Biochemistry* **1**, 144-149 (1988).
8. Penner, R., Matthews, G. & Neher, E. *Nature* **334**, 499 (1988).
9. Fink, L.A. & Kaczmarek, L.K. *Trends neurosci.* **11**, 338-339 (1988).

Glucose transport

SIR—Insulin is known to stimulate the production of diacylglycerol after binding to membrane receptors, and Strålfors has now shown¹ that exogenous diacylglycerols mimic insulin in stimulating glucose uptake into adipocytes. Based on this finding, he postulates that endogenous diacylglycerol mediates the effect of insulin on glucose transport. An important prediction of this hypothesis is that treatments which prevent the production of, or neutralize, diacylglycerol in the plasma membrane should be capable of blocking insulin-mediated glucose uptake. We believe that data published by us²⁻⁴ sup-

port this prediction, thereby lending credence to Strålfors' hypothesis.

In our studies, we examined the effect of the cyclic polycationic decapeptide polymyxin B on various actions of insulin. In our first experiments², we found that polymyxin B prevents insulin-induced hypoglycaemia in experimental animals. Subsequently, we observed³ that this compound blocks insulin-mediated 2-deoxyglucose uptake in rat adipocytes. Furthermore, we found⁴ that polymyxin B inhibits insulin-dependent glucose transport without affecting insulin binding to the receptor or altering the ability of insulin to either induce receptor autophosphorylation or activate receptor kinase activity. Other insulin-dependent actions, such as inhibition of lipolysis in adipocytes, synthesis of DNA in muscle cells or activation of glycogen synthase in mouse skeletal muscle are not affected by polymyxin^{3,4}.

Previous work had demonstrated that polymyxin B binds electrostatically to acidic phospholipids in membranes⁵, and it is possible, therefore, that the inhibitory effect of polymyxin B on insulin-stimulated glucose transport reported by us² is due to just such a peptide-phospholipid interaction, leading to neutralization of a membrane phospholipid. Indeed, acetylation or succinylation of the five amino groups of polymyxin B prevents both binding of the peptide to membrane phospholipids as well as its ability to block insulin action³. Thus our observations suggest the existence of a polymyxin B-sensitive membrane phospholipid that functions in a highly specific manner in the mechanism that mediates the effect of insulin on glucose transport. This phospholipid may be a diacylglycerol, as suggested by Strålfors.

SHIMON AMIR

Center for Studies in Behavioral
Neurobiology,
Concordia University,
Montreal, Quebec, H3G 1M8, Canada

YORAM SHECHTER

Department of Hormone Research,
The Weizmann Institute of Science,
Rehovot, Israel

1. Strålfors, P. *Nature* **335**, 554-556 (1988).
2. Amir, S. & Shechter, Y. *Eur. J. Pharmac.* **110**, 283-285 (1985).
3. Amir, S., Sasson, S., Kaiser, N., Meyerovitch, J. & Shechter, Y. *J. Biol. Chem.* **262**, 6663-6667 (1987).
4. Shechter, Y., Meyerovitch, J. & Amir, S. *Biochem. Pharmac.* **37**, 1891-1896 (1988).
5. Craig, W.A. & Kunin, C.M. *J. Pharmac. exp. Ther.* **184**, 757-765 (1973).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

The fifth nuclear power

Lawrence Badash & Margo Alexander

China Builds the Bomb. By John Wilson Lewis and Xue Litai. *Stanford University Press*: 1988. Pp. 329. \$29.50.

THE development of the Chinese atomic bomb has long been shrouded in mystery. Why did the People's Republic decide to obtain nuclear weapons? How was this done in a nation recovering from civil war and lacking in food, let alone high technology? To what extent did the Chinese receive help from abroad? In the past decade, numerous articles and three major books have been published in Chinese on various aspects of the building of the bomb; the work under review is the first synthesis in English of this literature.

Mao Zedong liberated China in 1949 by following his concept of a 'people's war'. His generals learned a different lesson in 1950–1951, when their troops were decimated by the greater firepower of US forces in Korea, and then felt particularly threatened by the US strategic policy of massive retaliation as tensions developed over the Taiwan Strait and then Vietnam. These external political crises drove the Chinese leadership to regard nuclear weapons of their own as a way to emerge from under the shadow cast by the United States.

More general factors were also involved in the decision to go nuclear. Nationalistic pride, and Mao's determination to create a militarily powerful state that would avoid the humiliations of the past, led in January 1955 to the Politburo decision to build strategic forces. There was a contradiction in this, for Mao distrusted technical experts. He retained his belief that masses of people determine the outcome of wars and called the atomic bomb a 'paper tiger'. But he was capable of allowing expediency to overrule ideology when the dignity of China was at stake.

The Soviet Union soon announced its intention of providing China with a cyclotron, a reactor, fissionable material for research and hundreds of scientific advisers, and of educating numerous Chinese students. Not so widely publicized was a commitment to help China build its bomb by allowing access to technical data, engineering drawings of plants and machines, apparatus, and even a prototype weapon. In return, China would provide Russia with raw materials. The Chinese appear to have had few illusions; they needed Soviet help, but intended to

be dependent on it for as short a time as possible. What is not clear is why the Kremlin was so obliging; peace pacts and fraternalism seem inadequate explanations.

In the book, the authors give overmuch attention to the wide search for uranium ore and the details of its extraction. Deposits were found in several provinces and, with great hardship, eight large mines were put into operation. This coincided with the Anti-Rightist (anti-intel-

lectual) Campaign of 1957, and the Great Leap Forward of 1958–1960, which left the country in so much turmoil. Such mass movements engaged China's greatest resource, its people, who in this event embarked on a vast uranium hunt. In something of a cottage industry, unskilled labourers performed tasks normally reserved for trained workers, and extracted enough uranium ore from shallow deposits (not the proper mines) in Hunan, Guangdong and Liaoning Provinces for the first bomb.

With Soviet assistance, probably including provision of the critical nickel barriers with their millions of microscopic holes, a gaseous diffusion plant to enrich uranium-235 to bomb grade purity was erected in the north-western Chinese province of Gansu, at Lanzhou. A plutonium-producing reactor, chemical extraction facilities, a plutonium metal fabrication plant, and a factory to convert Lanzhou's gaseous uranium hexafluoride product into uranium metal were built

near Subei, also in Gansu. The Chinese, of course, knew that the bomb could be built, but this did not relieve them of the need to scour the foreign literature, theoretically and experimentally develop many techniques, and overcome numerous design and production hurdles. They specialized in small-scale, trial-and-error work, constructed apparatus rather than spent foreign exchange on it, copied rather than created and were content with lower levels of sophistication — if they worked.

At first, neither plutonium nor uranium-235 was given priority. Work on the latter happened to be further advanced by 1960, when a decision on the material for the first bomb was made. Interestingly, the Chinese chose to design an implosion device, rather than the simpler gun type in which one piece of subcritical material is fired at another. Lewis and Litai recognize that the symmetrical compression of the fissionable material is a superior method, although they fail to explain why (less material is required for supercriticality).

China's equivalent to Los Alamos, the bomb design laboratory (the Northwest Nuclear Weapons Research and Design Academy, also called the Ninth Academy), was built in the wastes of Qinghai Province. It was led by a 43-year-old general, Li Jue, who had two decades of military, political and defence production experience behind him. Here were concentrated the élite scientists of the nation, several with doctorates from foreign universities: Zhu Guangya (nuclear physics, University of

Michigan), Wang Ganchang (nuclear physics with Lise Meitner, University of Berlin), Peng Huanwu (quantum physics with Max Born, Edinburgh University), Guo Yonghuai (aerodynamics, California Institute of Technology). As at Los Alamos, openness was encouraged inside the fence. All too little appears to be known about the scientists, of the technical problems they encountered and of the industrial contributions, but from the book we do learn, for example, that construction of the neutron initiators proved to be far more difficult than design and fabrication of the explosive lenses needed for implosion.

By 20 August 1964, the assembled bomb, less its enriched uranium-235 core, was ready for rail transport from Subei to the Lop Nur test site, in the Xinjiang Uygur Autonomous Region, a province in the far northwest of China. The core followed soon after by air. On 16 October 1964 (the apparent delay is unexplained), the bomb was successfully exploded atop a

CHINA EXPLODES HER ATOMIC BOMB

CALL FOR SUMMIT TO BAN NUCLEAR ARMS

DAVID FLOYD

Daily Telegraph Correspondent on Communist Affairs

CHINA exploded her first atomic bomb yesterday, somewhere in Western China. She thus joins America, Britain, Russia and France, as the world's fifth nuclear Power.

The official statement from Peking described the explosion as "a major achievement of the Chinese people" and "a triumph of the Chinese people's

Expediency over ideology — front page news on 17 October 1964.

tall metal tower, with a yield of about 20 kilotons, the same size as the first bombs of other nations.

This achievement is remarkable for, despite the real contributions made by the Soviets early in the programme, Moscow reneged on its promises of providing a prototype bomb and data and equipment, and withdrew all of its advisers by mid-1960. Hostility between Mao and Khrushchev surfaced: the Soviet leader felt that Mao would be reckless with nuclear weapons, while Mao berated Khrushchev for his adventurism and then capitulation in the Cuban missile crisis, and felt that the Soviets were drawing too close to the West by signing the Limited Test Ban Treaty and advocating peaceful coexistence. Moreover, the bomb project succeeded despite the frenzied years of the Anti-Rightist Campaign, the Great Leap Forward, the 'three hard years' (1960–1962) and the beginning of the Cultural Revolution.

The overall head of the Chinese bomb project, vice-chairman of the Central Military Commission Nie Rongzhen, appears as a highly competent administrator. Yet the confusing chop suey of institutes, academies, bureaux, commissions and ministries was not the key to success. Rather, flexibility in the face of changing politics, pragmatism and the old-boy network proved to be far more efficient — and necessary. Supremely important was the feeling of national pride, adventure, commitment and camaraderie among the scientists, technicians, civilians and military personnel that was engendered by the project.

We must express a feeling of unease regarding the accuracy and comprehensiveness of the information available to Lewis and Litai. At bottom, the data in the book are what the Chinese government wanted to release. Moreover, one of the three main sources relied upon, which we have seen (*Mimi Licheng*), is such a vague and anecdotal record that it must be used with caution. We do not know the quality of the other sources, but the authors' unrestrained inclination to quote and paraphrase prose in the 'heroic' style of revolutionary realism and romanticism is not expected in Western historical scholarship. Furthermore, most of the technical accounts are collages of quotations, which suggests that the authors had only a rudimentary comprehension of such matters. Yet, whatever the defects, this is the only full-scale book on the Chinese bomb in English and it must be consulted by those interested in the subject. □

Lawrence Badash is Professor of History of Science in the Department of History, and Margo Alexander is a Chinese scholar in the Asian Studies Program, University of California, Santa Barbara, California 93106, USA.

Spring in the step

Thomas A. McMahon

Elastic Mechanisms in Animal Movement.

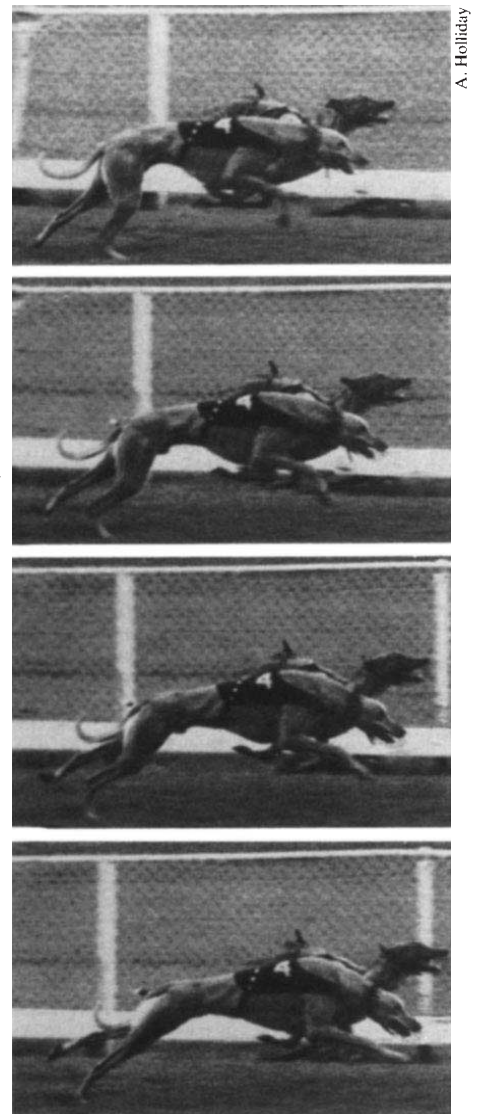
By R. McNeill Alexander. Cambridge University Press: 1988. Pp. 141. £19.50, \$34.50.

CONSIDER the following paradox. Studies are made of a kangaroo hopping and a man running. Both are filmed using high-speed ciné cameras, and both run over force platforms capable of measuring the ground reaction force. Calculations based on the measurements show that the energy lost and regained per metre travelled is more than can be supplied by the muscles alone, given the measured rate of oxygen consumption and a reasonable value for the energetic efficiency of muscle. The paradox is resolved only if it is assumed that somehow springs are used to store part of the kinetic energy the body had before landing. The springs then supply the missing energy required to explain the take-off at the end of the step.

Since Giovanni Cavagna and his colleagues first described this paradox in human runners and proposed its resolution in 1964, the search has been on for the site of the springs and an elucidation of the way they work. R. McNeill Alexander has been a leader in this search, and he has looked not only in men and kangaroos but in selected places throughout the entire animal kingdom for evidence of important springs. Now he writes about his own success stories and those of others in this charming, small volume *Elastic Mechanisms in Animal Movement*.

Employing the spare and clear style he used to great advantage in his classic *Animal Mechanics* (first published in 1968), Alexander first introduces a mechanical subject — for example, the suspension system of a car — and then applies the results to a zoological problem, in this case the design of a paw pad which won't 'chatter' when it is planted on the ground. His method is always to develop an understanding based on clear arguments and then introduce experiments designed to test the conclusions. For example, in discussing catapult mechanisms in insects, he presents a study done on jumping locusts. He focuses on the semilunar processes and the extensor apodeme of the hindleg, and concludes that the total energy that could be stored in each leg, 6 mJ, is more than the total energy required to power the jump (4.4 mJ per leg). Thus all the energy required for a jump could be stored before the jump by slowly winding up the elastic structures.

The accounts of zoological elastic mechanisms he presents are fresh and



A. Holliday

Power plus — energy is saved in the greyhound's gallop by the elasticity of tendons in the legs and back.

entertaining. I want to mention two of my favourites. One is the discussion about how the accelerations and decelerations of the running movement could drive breathing displacements in dogs and other quadrupeds. The other concerns the jumping of click beetles, and how the beetle jumps from a lying position by suddenly bending its back at a joint between two segments of the thorax. Alexander says this style of jumping leaves the beetle oscillating, bending and extending its back as it flies upward, so that about one-third of the energy used for jumping goes into the oscillations and is wasted. He points out that this is equivalent to an animal jumping with a heavy foot, and lets readers see exactly why, if the aim is to jump as high as possible, they should take their boots off first. □

Thomas A. McMahon is Gordon McKay Professor of Applied Mechanics and Professor of Biology, Pierce Hall, Harvard University, Cambridge, Massachusetts 02138, USA.

Look here upon this picture and on this

P.W. Hawkes

Image Processing, Continuous to Discrete. Vol. 1, Geometric Transform and Statistical Methods. By Edward R. Dougherty and Charles R. Giardina. *Prentice-Hall:1987.* Pp.452. \$78.65, £61.

Image Recovery: Theory and Application. Edited by Henry Stark. *Academic:1988.* Pp.543. \$75, £56.

Morphological Methods in Image and Signal Processing. By Charles R. Giardina and Edward R. Dougherty. *Prentice-Hall:1988.* Pp.321. \$78.65, £61.

Image Analysis and Mathematical Morphology. Vol. 2, Theoretical Advances. Edited by Jean Serra. *Academic:1988.* Pp.411. \$89.50, £44.

Matrix Structured Image Processing. By Edward R. Dougherty and Charles R. Giardina. *Prentice-Hall:1987.* Pp.258. \$75.90, £58.85.

Digital Pictures: Representation and Compression. By Arun N. Netravali and Barry G. Haskell. *Plenum:1988.* Pp.586. \$90, £52.95.

Optical Signal Processing. Edited by J. L. Horner. *Academic:1987.* Pp.569. \$72, £45.50.

THESE seven substantial volumes, between which there is some but not much overlap, illustrate both the extent and the disparate nature of the area that we call image processing. They also show that the subject has come of age, in the sense that several of them have the lasting quality of treatises, despite the fact that some branches of the subject are still developing (image analysis and description are obvious examples).

Traditionally, image processing is divided into four provinces: acquisition and coding; enhancement; restoration; and analysis and description, including pattern recognition. All of the books can be fitted into one or more of these categories, but in *Image Processing* Dougherty and Giardina take a different and more fertile viewpoint. Here, image processing is regarded as the sequence that takes us from image acquisition, or sensing, to making a decision about the detection of some structure; between the two may lie restoration, enhancement, segmentation, feature selection, registration and classification. The material is the same, but an objective is declared from the outset and the various operations become subservient to it.

Henry Stark's edited work, *Image Recovery*, is concerned exclusively with restoration, by both linear and nonlinear methods. It contains chapters on the inverse problem in general, statistical methods of dealing with incomplete data, linear programming and maximum entropy; several are devoted to various aspects of the phase problem and of tomography. The contributors are distinguished, many of them having made notable contributions to the subject, with the result that this collection is a particularly useful and authoritative addition to the literature.

The three volumes by Dougherty and Giardina are written at very different levels. *Image Processing* is the first of a series of introductory texts, and covers only a small part of the themes mentioned

above. In the first of the six main chapters, the authors explain in some detail matrix representation of images, including their general notion of bound matrix within which any, not necessarily rectangular, array of image values can be embedded. An account of mathematical morphology follows, in which the notation that newcomers find highly indigestible is seasoned with worked examples. Then come two very full chapters on transform techniques, in which I feel the authors go into somewhat wearisome detail — the book is, after all, aimed at the "professional image engineer" and his undergraduate courses will have covered this material (though possibly not this application) thoroughly. Incidentally, Dougherty and Giardina fail to define the Hough transform, in the section devoted to it, and unwisely only give examples; indeed an evil spell seems to have been cast on this transform, for the associated figure on p. 239 has vanished. The penultimate chapter presents the statistical theory needed in image processing, while the final chapter deals with statistical estimation.

The book is original in a way that should appeal to readers whose training has been dominated by computer programming, for much of the mathematics is presented in a hybrid form, half traditional, half computer instruction. Almost every operation is illustrated with an example, and although the scope is limited this is a book which both teachers and students will find very helpful. It does not, however, supplant the magisterial *Digital Picture Processing* by A. Rosenfeld and A.C. Kak, published by Academic in 1982.

Of the other two books by Giardina and Dougherty, *Morphological Methods* is an amplified version of the chapter on mathematical morphology in their more elementary text, and is to be compared with the volume edited by Serra, a successor to his *Image Analysis and Mathematical Morphology* (Academic, 1982). The literature of this subject is expressed in the language and notation of sets and abstract

logic, with the result that many potential users are rebuffed. Both Giardina and Dougherty, and the contributors to Serra's book, are aware that this notation can be rebarbative and one feels, as one reads the texts, the seductive force of the mathematical formalism, so well adapted once one has become accustomed to it, straining against the realization that readers are very likely to give up before they have become accustomed to it.

In both books, many examples are given and the morphological techniques are related to concepts, such as the skeleton, which students of image processing will have encountered already. Although both are hard going, the new Serra is a great deal easier to assimilate than its predecessor and when we have Vol. 3, on algorithms, the set will surely form the ultimate authority on the subject. Nevertheless, I encourage anyone determined to master the subject to acquire Giardina and Dougherty as well, for they have taken great pains to be comprehensible and often shed light on difficult passages in Serra. Specifically, they open with chapters on morphology in the euclidean plane and on digital morphology, with careful accounts of Minkowski algebra and of opening and closing, dilation and erosion. A third chapter introduces the idea of morphological features, and examines image functionals, granulometry and digital size-distribution analysis. This leads on naturally to a discussion of topological processing, where we meet the skeleton and Serra's hit-and-miss operator. Three other chapters deal with morphological filters, for binary and grey-scale images, and there is also an account of how the earlier notions are extended to cover images with multiple grey levels.

Of the 18 chapters of Serra's book, 15 are by him or G. Matheron and between them they describe in great detail all the themes of this subject, of which they are the creators: morphology for complete and for boolean lattices, dilations on topological spaces and structuring functions. Six chapters cover various aspects of filters, three others skeletons. A chapter explains how measurements are made on numerical functions. Two others analyse convexity and symmetry. Finally, there are accounts of the important theme of boolean random functions and their use in boolean texture analysis and synthesis.

The other book by Dougherty and Giardina, *Matrix Structured Image Processing*, is a relatively elementary introduction to image processing, many of the chapters being simpler versions of their image processing text. The notation is similar and a good (human) memory is essential to remember exactly what the numerous operators do (there are over a hundred in the index). The seven chapters introduce the newcomer to fundamental operations, grey-level manipulation, edge

From Serra's *Image Analysis*

Now you see it . . . The computer image of a cuboctahedron (left) has been deliberately buried by white noise (centre) to show that an algorithm based on 'morphological filters' can uncover the hidden form (right). Paradoxically, the routine uses operations that lose information in retrieving the ordered structure, even though noise represents the converse of information.

detection, morphology, topological operations, transforms and imaging architectures, notably systolic and wavefront array processors.

Netravali and Haskell's *Digital Pictures* is concerned entirely with coding, which is important for different reasons: images have to be transmitted, and it is desirable to do so as efficiently as possible, while if images are archived they must be stored as compactly as possible without undue loss of information.

After dealing with the ways of representing visual information numerically and with the various picture communication systems, the authors provide a long account of the various types of code that permit image data to be compressed. First they describe codes that exploit redundancy, of which the Huffman codes and those based on discrete transforms are typical examples. Luminance statistics, so important for the predictive techniques, are then analysed and the usual models introduced. These ideas are then extended to include colour pictures and binary graphics. An important chapter on visual psychophysics follows, indispensable for analysing the decoding step.

Perhaps the true heart of this book, however, is the 200-page chapter on basic compression techniques in which all the important methods are presented (though I should have liked to see more space devoted to vector quantization). Few authors can have succeeded so well in practising what they preach, for Netravali and Haskell have compressed a vast amount of data into a single volume and the human task of decoding it is not unduly daunting.

The final book, *Optical Signal Processing*, edited by J. L. Horner, clearly belongs here for its chapters have much the same titles — pattern recognition, transformations and signal processing — that are found in image-processing texts, but it differs from all the others in that the processing is performed optically, not electronically. Optical techniques are attractive because large numbers of pixels can be manipulated quite easily, whereas most digital image processing is performed on rather small image matrices. There are, of course, disadvantages too, as the editor

cheerfully acknowledges: "... the birth of the laser and the discovery of off-axis holography in the early 1960s got the field off to a running start. In the intervening years, the field has seen several cycles of bloom and doom".

The book is divided, rather haphazardly, into seven broad sections, each with several contributors, and sadly there is no introductory chapter by the editor, setting the scene and imposing some kind of pattern on all this material. There are sections on white-light processors, pattern recognition, temporal processing and nonlinear processors, many of which are written by respected authorities but which all too often lack the introductory material that would have made them so much more accessible. The next section, however, on transformations, is easier to penetrate; in one part, tomographic transformations are considered, while in the other we are told how the various transformations needed in image processing are performed.

As we approach the end of the book, so the authors come closer to the present and venture into the future. The section on optical numerical processing includes a lively chapter on linear-algebra processors, essentially on ways of manipulating matrices and vectors, and a properly cautious and minatory account of algorithms and software by H. J. Caulfield: "[Optics can] offer truly massive parallelism, say 10^{12} channels, which is achievable in electronics only with equally massive inconvenience", he observes, before pointing out that, as yet, there is no convenient way of performing floating-point operations optically and that optics will not develop a high-level language until standard architectures and standard routines have emerged. This short chapter gives no more than the flavour of things to come.

In Horner's book we thus have a good account of work in progress by those most intimately concerned with pushing back the frontiers of optical processing. But the editor would surely agree with Caulfield's observation that "No text of lasting significance is possible at this stage". □

P.W. Hawkes is *Maitre de Recherches* at the *Laboratoire d'Optique Electronique du CNRS*, BP 4347, 31055 Toulouse Cedex, France.

Desert island dictionary

T. A. Connors

Macmillan Dictionary of Toxicology. By Ernest Hodgson, Richard B. Mailman and Janice E. Chambers. *Macmillan*, London:1988. Pp.395. £39.50. In the United States available from Van Nostrand Reinhold, \$67.95.

TOXICOLOGISTS work on an immense variety of topics. In consequence those involved in basic research may often find themselves needing to understand terms which are not immediately related to their own area but which are important in solving a particular problem. Acronyms such as GRAS (generally regarded as safe) or TLV (threshold limit value) are familiar but not always fully understood. On the other hand, those involved in safety evaluation need to understand the precise meaning of, for example, bioavailability or covalent binding. From time to time all of us come across the formulae of chemicals normally referred to by their trivial names, while students and newcomers to the subject have a clear need for guidance in getting to grips with its basic concepts.

This dictionary, another one in the Macmillan series, is a *tour de force* in that most of these diverse requirements are met in a compact volume of less than 400 pages. The entries are generally precise and accurate; they include the most frequently used toxicological terms and definitions, and incorporate clearly drawn formulae of the most important chemicals.

The authors specifically invite both corrections and additions from readers for the next edition. I would suggest the inclusion of anti-oxidant and the sharpening up of some definitions, for instance that of asbestos where there ought to be some distinction between blue asbestos (crocidolite) linked to mesothelioma in humans and the other fibres associated with asbestosis and increased lung cancer risk especially in cigarette smokers. And the cytochrome P-450 cycle, where there is a terrible mix-up in the electronic structure of some of the intermediates, should certainly be corrected.

Scientists stranded on a desert island and given a limited choice of books would probably select good reference works. My personal choice would be the *Merck Index*, *Documenta Geigy* (before it was split into sections) and, of course, Goodman and Gilman. Toxicologists anticipating shipwreck could well consider adding this dictionary to their lists. □

T.A. Connors is *Director of the MRC Toxicology Unit, Medical Research Council Laboratories, Woodmansterne Road, Carshalton, Surrey SM5 4EF, UK.*

Allocation of resources for conservation

N. Leader-Williams and S.D. Albon

Theory dictates that conservation areas should be as large as possible. When money for their protection is inadequate, different considerations come into play.

IN THEORY, large nature reserves or parks reduce the risk of extinctions because they contain sizeable populations of endangered species of plants and animals. In practice, however, most developing countries do not have the resources to protect large areas and economically valuable species from illegal exploitation. Here we show that the rates of decline of rhinos and elephants are related directly to conservation effort and spending. The conclusion to be drawn from these results is either that conservation schemes must be adequately funded, or resources must be concentrated in small parts of large reserves, if local extinctions are to be avoided.

Since the 1930s, many developed and developing countries have established national parks and nature reserves. The aim is usually to protect ecosystems or large parts of them, together with their indigenous floras and faunas, in a state relatively untouched by human exploitation or occupation. According to conservation biology theory, large protected areas minimize the risk of extinctions arising from genetic isolation because they contain sizeable populations, corridors between parks further reducing isolation^{1,2}. Several large conservation areas in Africa, such as Serengeti, Tsavo, Selous and Luangwa, come close to fulfilling these theoretical ideals, yet the recent declines of large populations of both black rhino (*Diceros bicornis*) and elephant (*Loxodonta africana*) within these areas^{3,4} shows that there is a wide gap between theory and reality. In this article we examine the reason why this should be so — first, by evaluating a management strategy that failed to protect black rhinos and elephants from illegal exploitation in Luangwa Valley in north-east Zambia (Fig. 1); and, secondly, by considering the wider implications for the funding of conservation where resources are limited.

The socio-economic background of protected areas is relevant when looking at conservation schemes; although there are local variations, the example of Luangwa Valley^{5,6} is fairly typical of many protected areas in Africa. The four national parks, totalling 16,660 km² in area, were originally established as game reserves in the colonial era. The human inhabitants, who had previously used the areas' products both for subsistence (meat, firewood, honey) and trade (ivory, rhino

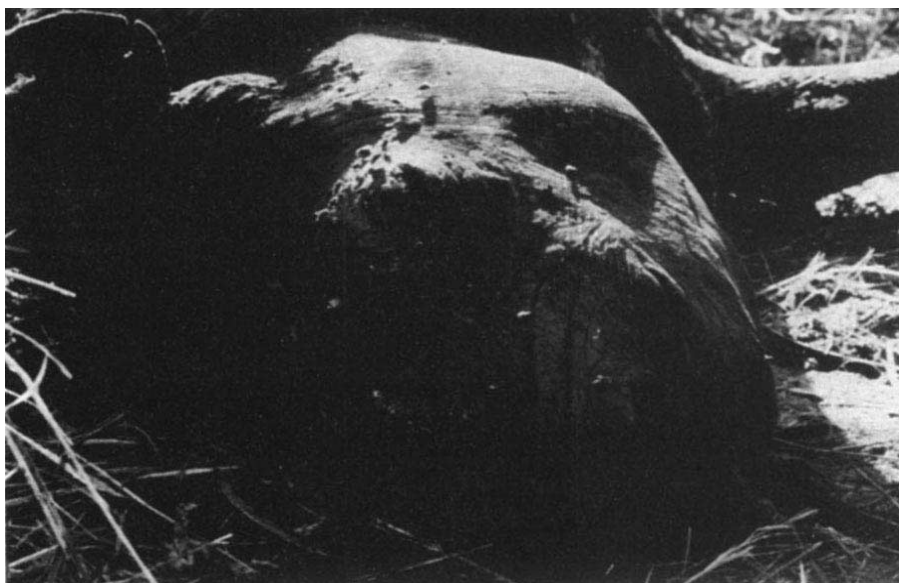
horn), were evicted. People were allowed to remain in seven sparsely inhabited hunting areas, totalling 46,342 km² that border on the reserves, but were subjected to game and gun laws and to licence quotas set to protect wildlife. Hence, both reserves and hunting areas were managed increasingly for the benefit of outsiders, chiefly tourists and safari hunters, and earnings from wildlife went largely to central government and the private sector. Local residents, denied access to resources that were previously under their control, became increasingly impoverished and resentful.

In 1972 Zambia affirmed its commitment to conservation by gazetting 9% of its surface area as national parks and 22% as game management areas. At that time, Luangwa Valley held large populations of elephant (100,000) and black rhino (between 4,000 and 12,000). However, Zambia's economy then began to decline because of falling copper prices, and although central government spent quite heavily on conservation, the amount was low in relation to the vast areas under protection. Consequently, park infrastructure and law-enforcement efforts began to collapse. By the late 1970s, Zambia's internal socio-economic problems, coupled with dramatic price increases of ivory and rhino horn on the world market^{3,4}, had resulted in a serious outbreak of poaching in Luangwa Valley.

By the mid-1980s elephants were reduced in numbers by 75% to around 25,000 and rhinos to probably a few hundred^{3,4}. Profits from this slaughter went not to the Zambian mainstream economy, but elsewhere — the smaller share to organized gangs who killed and extracted ivory and horn from animals within the parks, and the larger share, including foreign exchange, to middlemen who smuggled the trophies out of Zambia. The slaughter provided little benefit to Luangwa residents, because most of the poachers came from areas bordering onto, but outside, Luangwa Valley.

In late 1979 an anti-poaching operation, funded in part by the Zambian government, was set up. The following year an external agency donated a relatively large sum in conservation terms (half a million US dollars over three years) to the operation. The government bought vehicles and mobilized staff into units that undertook regular foot patrols in important areas with the aims of arresting poachers and protecting rhinos and elephants⁷.

In any such project, it is essential to monitor the numbers of animals under protection. We used an index of rhino abundance based on sightings made by patrols between 1979 and 1985 (Fig. 2). The reliability of the index was confirmed by comparing the results with a very accurate method of counting rhinos in one area⁸ (instantaneous rates of decrease of



N. Leader-Williams

A victim of poachers, its horn hacked off, is left to rot.

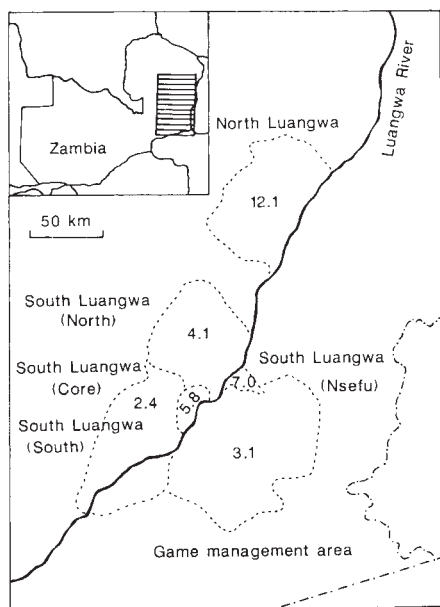


Fig. 1 Map of Luangwa Valley showing North and South Luangwa National Parks, the different sizes of six areas patrolled regularly and the frequency of rhino sightings per patrol in 1980 (see Fig. 2 for methods of assessment).

24% yr^{-1} against 29% yr^{-1}). Sightings of elephants (which are not depicted in this article) were monitored by patrols in the same way and produced similar rates of change to those obtained by aerial counts. Each area patrolled was of a different size, and patrol intensity varied between areas. At one extreme, the remote 4,636 km^2 North Luangwa National Park was visited by only 30 patrols totalling 322 patrol days. In contrast, a small area of 400 km^2 was designated as the core of South Luangwa National Park, and one or two patrols were permanently policing it from 1982, resulting in a total of 337 patrols (2,260 patrol days).

In spite of this protection, rhinos declined at an overall rate of 63% yr^{-1} throughout Luangwa Valley. However, their rates of decrease differed between areas, and varied between 24% yr^{-1} and 99% yr^{-1} (Fig. 2). Such high rates of decrease in a species with a maximum recruitment rate of 7–11% yr^{-1} resulted in the virtual disappearance of rhinos from most areas of Luangwa Valley between 1979 and 1985. By 1985 rhinos were seen regularly only in the two smallest areas, at frequencies of 1.78 and 1.04 rhinos per seven-day patrol (Fig. 2). Elephant numbers also declined at an overall rate of 12% yr^{-1} throughout Luangwa Valley. In one area they decreased at a rate of 42% yr^{-1} but in others, where there were fewer elephants to start with, their numbers actually increased at rates of up to 18% yr^{-1} . Elephants, however, are known to recognize areas of relative safety, and the increases probably reflect immigration rather than increased recruitment.

It is intuitively obvious — but infrequently demonstrated⁷ — that the resources put into a conservation scheme will relate directly to its ultimate success. In Luangwa Valley there was a direct relationship between the rate of decrease of rhino and elephant numbers in each area and the patrol effort, corrected for both size of area and frequency of rhino and elephant sightings in 1980 (Fig. 3). Hence within Luangwa Valley the patrols did deter poaching, but the effort was too thinly spread, even in the core area, to prevent the decline of rhinos. Extrapolation of the linear regression for rhinos predicts that a decline in numbers could have been prevented if five separate patrols had permanently covered the core area. But, given the available manpower, to have achieved this would have meant the concentration of all patrols in the core area, leaving 98% of the total area unprotected⁷.

A number of factors, such as external demand for trophies, poverty and corruption, result in poaching of rhinos and elephants. Given this, the detailed results from Luangwa Valley support the view that an important factor in the overall decline in rhino and elephant numbers across the rest of Africa is a shortage of manpower, and ultimately of resources, within national conservation departments^{10,11}. Both black rhinos and elephants are widely distributed across Africa, often sympatrically but with rhinos always at lower densities than the more adaptable and wide-ranging elephant. Surveys during the 1980s provide information on gross population trends of both species in different countries^{3,4}. As in Zambia, most large populations of rhinos and several large populations of elephants suffered serious declines. Rhino numbers remained the same or increased in only three countries, whereas elephant numbers increased in perhaps four. In 1980 a survey of manpower and spending by central government within conservation areas showed wide national differences^{10,11}, and there is a direct relationship between spending (corrected for total area) and the estimated changes in rhino (Fig. 4) and elephant numbers in each country. To achieve a zero decline of rhinos the regression predicts that spending should have been US\$230 $\text{km}^{-2} \text{yr}^{-1}$ (Fig. 4), and that \$215 was necessary for elephants. Rhinos normally occur at densities of 0.4 km^{-2} — if all protection costs are attributed to them as the most sensitive indicator species, this was equivalent to \$575 per rhino per year. In Luangwa Valley, with government spending of \$11 $\text{km}^{-2} \text{yr}^{-1}$, the external donation given to Zambia should have been spent entirely within 725 km^2 over three years, confirming that all patrol effort should have been devoted to a relatively small area that would have held

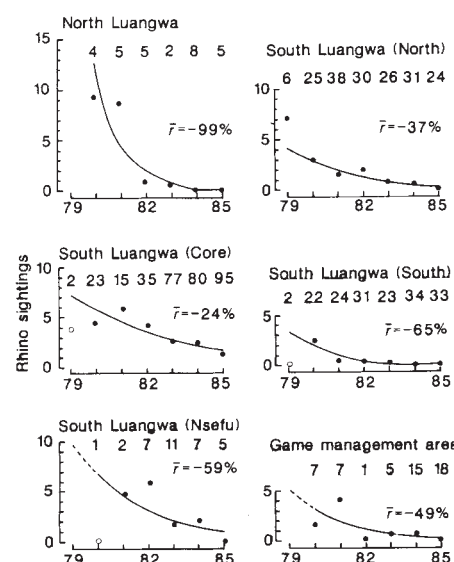


Fig. 2 Rates of decline of black rhinos in six different areas of Luangwa Valley, quantified from an index of abundance based on sightings made by anti-poaching patrols¹¹. A total of 768 patrols of different duration were undertaken in all months of the year between 1979 and 1985. These data were heavily skewed due to the absence of rhino sightings on many patrols, and therefore were analysed using log-linear models with a Poisson error distribution. Patrol length, number of scouts, month, area, year and an area-year interaction all had a significant effect on the basic model and considerably reduced its deviance. To take account of different patrol size and duration, and of season, all sightings of rhino were adjusted in the model as if all patrols had been made by four scouts over a seven-day period in January. For reasons of clarity, values are depicted as adjusted means for each year in each area (sample sizes shown above year values), but curves were fitted through all data points. Instantaneous rates of decline, assessed by including year as a linear variable, showed a highly significant difference between the six areas. Some areas were not patrolled, or patrolled only lightly in 1979 or 1980, so rates of decline were calculated from the first year with >2 patrols, thus omitting open symbols. Dashed lines show extrapolations based on rates of decline from 1980 or 1981, as appropriate.

almost 300 rhinos. But because the external donation was spread over approximately 16,000 km^2 , it added only a further \$10 $\text{km}^{-2} \text{yr}^{-1}$, totalling less than 10% of the spending necessary to prevent a decline of rhinos.

This study is of wide relevance to conservation practice. Conservation biologists have concentrated upon theoretical implications of the size of protected areas and of reduced population size and inbreeding depression^{3,12}, but there are few empirical studies that evaluate the efficiency of resource use in achieving conservation objectives. Because of limited funding^{10,11}, theoretical aspects of conservation biology may sometimes obscure realistic goals for developing countries. Our results show that the rates of a species' decline are related directly to protection

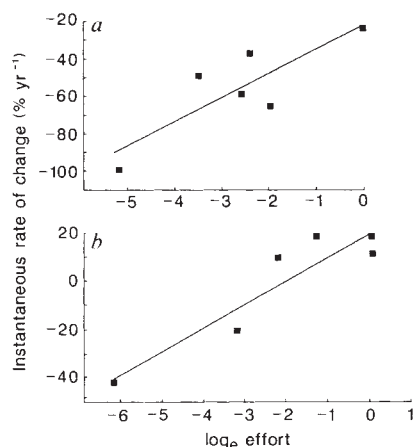


Fig. 3 *a*, Relationship between the instantaneous rate of decline of black rhinos (from Fig. 2) and patrol effort in different areas of Luangwa Valley. Because each area was of a different size and initially held different numbers of rhinos (see Figs 1 and 2), the effort for each area was calculated as log_e (total patrol days/size of area/frequency of rhino sightings in 1980). The regression equation was of the form $y = 0.13x - 0.16$ ($r^2 = 0.73$, $P < 0.05$). *b*, A similar relationship held for rates of change in elephant numbers and patrol effort, and the regression equation was of the form $y = 0.10x + 0.20$ ($r^2 = 0.83$, $P < 0.01$).

effort. If funding for conservation cannot be increased, then concentrating resources upon selected areas provides a pragmatic option for black rhinos and elephants, as well as for other endangered species, such as lowland gorillas or pygmy chimpanzees^{13,14}. Indeed, the success of sanctuaries in stabilizing numbers of the few remaining black rhinos in Kenya in the past year or so, and the recovery from near extinction of the southern white rhino and the vicuña, has resulted from pursuit of such policies¹⁵.

The relatively small sums that international conservation agencies have available to spend on valuable species in developing countries are most likely to achieve results in one of two contrasting situations — first, in low-spending countries only if they are concentrated at appropriate levels over small areas, such as the Kenyan rhino sanctuaries or the Virunga mountain gorilla population; secondly, over large areas only if the money is allocated to a relatively high-spending country like Zimbabwe, which is now in need of extra resources to prevent Zambians killing rhinos in the Zambezi Valley.

In poor countries, large conservation areas and sizeable populations of valuable species can probably only be maintained by a radical change in approach, combining the rectification of socio-economic problems^{6,15} with more investment in park infrastructure and policing. The most realistic option lies with appropriately directed conservation and rural development projects funded by international aid agen-

cies, either directly or through debt-swap schemes. In particular, residents of the areas concerned must be allowed to participate in plans for their local conservation areas¹⁶. If wildlife populations then recover, development can aim to become self-sustaining. For this to happen, earnings from sources such as tourism, licence fees and trophy sales must in part be recirculated, rather than going only to central government and business, both to pay for continued policing and to benefit local residents.

Such a project is now underway in Luangwa Valley¹⁷, with the elephant as its linchpin. It is worth bearing in mind that at least 75,000 elephants have been lost from Luangwa since the 1970s. A ban on the licensed hunting of elephants was imposed in 1980 in response to the escalating poaching, yet the sale of between 100 and 200 licences to foreign hunters between 1980 and 1985 would have raised a sum equivalent to the external donation given to Zambia in 1980, without even considering the value of the ivory or of the meat for local people. The aim, therefore, must be to achieve more balanced accounting, to develop the area and its wildlife for the benefit of the human inhabitants and to replace the current conflict between them and protected areas with a custodial and participatory relationship that benefits both parties^{5,6,17}. After the failure of underfunded protectionist policies in preventing serious declines in rhino and elephant numbers across Africa, schemes such as that now operating in Luangwa provide the best hope for the recovery of depleted populations and maintaining sizeable populations of valuable species in large conservation areas^{1,2,12}.

Our results demonstrate the principle that adequate resources need to be invested to achieve given objectives in conservation, whether for protection or development. Individual nations, funding agencies and conservation biologists together will have to determine policies which define how much conservation is enough and affordable, and make selective decisions instead of bowing to particular interest groups that believe all conservation is necessary. In determining how much conservation is affordable, the economic value of each species must be taken into consideration. Although

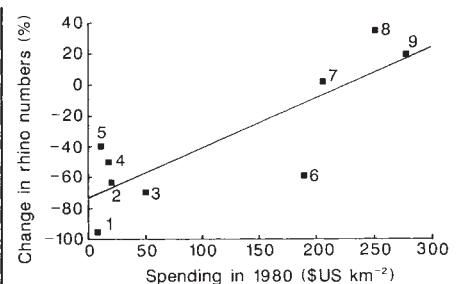


Fig. 4 Relationship between change in black rhino numbers between 1980 and 1984, and conservation spending in various African countries. Surveys of rhino numbers¹ were conducted in 1980 and 1984 and the percentage change was calculated for countries with > 250 rhinos in 1980 from these estimates as $(1980 - 1984/1980) \times 100$. Because the change in Zambia's rhinos was calculated with these independent data rather than from Fig. 2, Zambia is included in this figure. Details of spending per square kilometre by central governments on their conservation areas was available for nine of these countries^{9,10}. 1 = Central African Republic, 2 = Tanzania, 3 = Somalia, 4 = Mozambique, 5 = Zambia, 6 = Kenya, 7 = South Africa, 8 = Namibia, 9 = Zimbabwe. The regression equation is of the form $y = 0.32x - 72.82$ ($r^2 = 0.68$, $P < 0.01$) and at 0% change had an intercept of \$230 km⁻². The census data for large populations (> 1,000) of elephants between 1981 and 1987 are of variable quality³ but a similar relationship between changes in elephant numbers and spending appeared to hold for 14 countries. The relationship, which is not shown, is of the form $y = 0.24x - 51.97$ ($r^2 = 0.32$, $P < 0.05$) and at 0% change had an intercept of \$215 km⁻².

less valuable species or ecosystems may cost less to conserve, the required sum for rhinos and elephants will no doubt be higher today because of inflation and raised stakes. Both species are now more scarce, and ivory and rhino horn prices are still rising. An economic arms race is on and Africa — its people and its wildlife — will suffer yet further if it is not defused.

We thank the National Parks and Wildlife Services and commercial companies in Zambia, together with the People's Trust for Endangered Species, the Fauna and Flora Preservation Society and the New York Zoological Society, for supporting this research. Staff of Save the Rhino Trust, funded by the World Wildlife Fund, collected data on patrols, and colleagues in Cambridge helped with the manuscript (written whilst NLW held a Leverhulme Research Fellowship).

N. Leader-Williams and S.D. Albon are in the Large Animal Research Group, Department of Zoology, 34A Storey's Way, Cambridge CB3 0DT, UK.

1. Soulé, M.E. & Wilcox, B.A. (eds) *Conservation Biology: An Evolutionary-Ecological Perspective* (Sinauer, Sunderland, Massachusetts, 1980).
2. Frankel, O.H. & Soulé, M.E. *Conservation and Evolution* (Cambridge University Press, 1981).
3. Douglas-Hamilton, I. *Oryx* **21**, 11–24 (1987).
4. Western, D. & Vigne, L. *Oryx* **19**, 215–220 (1985).
5. Marks, S.A. *The Imperial Lion: Human Dimensions of Wildlife Management in Central Africa* (Westview, Boulder, Colorado, 1984).
6. Abel, N. & Blaikie, P. *Envir. Manage.* **10**, 735–751 (1986).
7. Leader-Williams, N. *Oryx* **19**, 27–33 (1985).
8. Leader-Williams, N. *Afr. J. Ecol.* **26**, 181–187 (1988).
9. Harcourt, A.H. *et al. Afr. J. Ecol.* **21**, 139–142 (1983).
10. Cumming, D.H.M., Martin, R.B. & Taylor, R.D. in *The Status and Conservation of Africa's Elephants and Rhinos* (eds Cumming, D.H.M. & Jackson, P.) 46–62 (Internat-

11. Bell, R.H.V. & McShane-Caluzi, E. (eds) *Conservation and Wildlife Management in Africa* (Peacocks, Washington, 1986).
12. Soulé, M. (ed.) *Viable Populations for Conservation* (Cambridge University Press, 1986).
13. Tutin, C.E.G. & Fernandez, M. *Am. J. Primat.* **6**, 313–336 (1984).
14. Susman, R.L. & Mubalamata, K.K. *IUCN/SSC Primate Specialist Group Newsletter* **4**, 34–36 (1984).
15. Jewell, P.A. & Holt, S.J. (eds) *Problems in Management of Locally Abundant Wild Mammals* (Academic, London, 1980).
16. Anderson, D. & Grove, R. (eds) *Conservation in Africa: People, Policies and Practice* (Cambridge University Press, 1987).
17. Dalal-Clayton, D.B. *Gatekeeper Papers* **9**, 1–14 (1988).

Spin and non-locality in quantum mechanics

C. Dewdney*, P. R. Holland†, A. Kyprianidis† & J. P. Vigiér†

* Department of Applied Physics, Portsmouth Polytechnic, Park Building, King Henry I Street, Portsmouth PO1 2DZ, UK

† Laboratoire de Physique Théorique, Institut Henri Poincaré, 11 rue P. et M. Curie 75231 Paris Cedex 05, France

Spin superposition in neutron interferometry, spin measurement, and non-local Einstein–Podolsky–Rosen spin correlations can be understood in terms of well-defined individual particle trajectories with continuously variable spin vectors.

EVER since its inception, various interpretations of quantum theory have been proposed and debated in the physics community in an attempt to give some physical meaning to its mathematical structure. The debate continues unabated to this day¹. What is at stake is not so much the consistency of the formalism and hence the predictive ability of the theory, but the nature of the reality that it describes.

What has come to be accepted as the orthodox view was set out by Bohr². Bohr saw in quantum theory a concise expression of his philosophy of complementarity, which he developed through a consideration of the limitations of our means of expression in various circumstances. He argued that the scientific enterprise must be based on unambiguous communication. In particular, the results of experiments must be capable of unambiguous description. This is taken for granted in classical physics, where the observer plays no special role and measurement is simply an ordinary physical process; there is no problem in conceptually distinguishing the observed system from the measuring apparatus. The classical conceptual scheme and its application in the realm of macroscopic measuring devices is not considered problematic by Bohr. Indeed he insists that all communication must be within its framework if it is to be meaningful. Bohr argued that “However far the phenomena transcend the scope of classical physical explanation the account of all evidence must be expressed in classical terms”². This injunction has profound consequences in the quantum domain, wherein the indivisibility of the quantum of action makes it impossible to separate the system undergoing observation and the observing apparatus. They form an unanalysable whole. The fundamental object of description in quantum mechanics is just this closed phenomenon, consisting of system and apparatus, and it makes no sense to attribute properties to individual quantum systems. Bohr argues not that our classical concepts are deficient, but that we need to be very careful about how we apply them. Through a process he describes as natural generalization, the conditions under which the classical concepts apply are to be delimited. In this way the concepts of particle and wave must be considered complementary. They require mutually exclusive experimental arrangements for their unambiguous application. There is no single concept or set of concepts that apply in all situations to allow us to represent an individual system precisely. The appropriate interpretation of the wave function is only (through the square of its modulus) as that which determines the statistical predictions appearing under conditions defined by classical physical concepts. The wavefunction represents the most complete possible specification of the state of an individual system. Bohr concluded that it is impossible to find any unambiguous language in terms of which one could discuss the origin of the statistical distribution of the results of experiments predicted by quantum theory. Indeed he states explicitly that “there is no question of reverting to a mode of description that fulfills to a higher degree the accustomed demands regarding pictorial representation of the relation between cause and effect”².

Although Bohr’s approach is consistent with quantum mechanics, it is not without its difficulties. Bohr proposes that the problematic features of quantum theory are resolved by the adoption of a particular epistemology (akin to the positivist position), in which the classical level of experience must be taken as a prescription, rather than derived from the theory. There is nothing in the bare quantum formalism and its statistical significance that compels us to adopt Bohr’s interpretation. We need not give up the idea of ascribing to an elementary system well-defined and continuously variable co-ordinates. Bohr’s position simply represents one possible point of view.

That there is another point of view with as long a history and an equal pedigree is not widely known. A radically different approach to Bohr’s was proposed originally by de Broglie in the 1920s^{3,4}. Although the experimental content is the same, the contrast between the way the two approaches accommodate what is novel in quantum theory is quite stark. When de Broglie proposed that microparticles should have wavelike characteristics, he had in mind that the associated wave is physically real, and not simply a mathematical artefact for computing probabilities. In this model, particles have continuous spacetime trajectories with well-defined momenta, as in classical physics. The wave-like distribution of trajectories is brought about by the “quantum potential” which depends on the form of the associated wave. Once the initial wave function and the initial position of the particle is given, its motion is uniquely determined through the evolution of the wave, according to Schrödinger’s equation. One set of concepts applies to all phenomena involving a given system, allowing them to be comprehended within a single picture. Instead of proposing a particular epistemology, this approach proposes a particular ontology, in which the classical level emerges in a natural way under certain circumstances defined by the theory.

This approach was later systematically developed, in the non-relativistic case, by Bohm⁵, who showed that it leads to the usual experimental predictions. In the de Broglie–Bohm approach, the observer plays no special role; indeed measurement is not a fundamental axiom of the theory postulated at the outset, but rather an example of a physical process in which two systems become correlated. The existence of the classical level is not assumed *a priori* but has to be derived as a limiting case of quantum theory. In this very different point of view we may picture individual quantum systems in a clear and unambiguous way, and we can discuss their interactions with other systems, some of which will be measuring apparatuses, in terms of a single conceptual scheme. The indivisibility of the quantum of action, the wholeness of the system and the apparatus, or more generally the system and its context, is expressed through the quantum potential. Heisenberg’s uncertainty relations are interpreted as statistical scatter relations. The primary role of the wave function is a physical one, determining particle dynamics, and its statistical significance is a derived consequence. In this way the properties of an ensemble of particles associated with the same wavefunction depend on the properties

of the individual and not the other way round. The necessity for a statistical treatment arises from the practical limitations on the control of the actual initial co-ordinates of the particle, and not from an inherent indeterminism in the properties of matter. As much as in classical physics, unique initial conditions lead to unique outcomes, through the operation of a set of causal laws.

Because it is able to discuss individual processes, the de Broglie-Bohm or causal theory also makes predictions which go beyond those possible in Bohr's approach; that is, it says what will happen to a particular particle in a given context. This has been illustrated for several typical quantum situations: two-slit interference⁶, tunnelling⁷, the Aharonov-Bohm effect⁸, neutron interferometry⁹ and two-body phenomena^{10,11}. Naturally these more detailed predictions cannot be experimentally verified so long as one is constrained by experiments carried out in the domain to which quantum mechanics applies.

The two points of view discussed here clearly approach the novel content of quantum mechanics in completely different ways. For Bohr, the important lesson concerns what we can say about nature; for de Broglie and Bohm, it concerns the way nature is.

The examples mentioned above, for which the causal de Broglie-Bohm theory has been worked out in detail, are all integral-spin, or bosonic systems. For such systems, a straightforward correspondence exists between the quantum particles and a classical wave equation. What is not so well-known is that one can consistently extend the causal theory to embrace spin- $\frac{1}{2}$, fermionic systems. Indeed, the claim that it is impossible to construct a realistic quantum model of matter is often thought to be proved for systems with spin, which exhibit a characteristic non-classical 'two-valuedness'. Nevertheless, just such an account has been developed by Bohm *et al.*¹², Takabayashi¹³ and more recently by us¹⁴⁻¹⁸. This is achieved by ascribing new properties to the wave, which is a spinor satisfying the Pauli equation and containing a spin-dependent addition to the quantum potential, and to the particle, which possesses a continuously variable spin-vector describing internal degrees of freedom in addition to the usual position and momentum variables.

Here we review recent developments in the causal theory of spin- $\frac{1}{2}$ particles which illustrate in terms of explicit graphics the one and two-body Pauli theory. The striking features which arise when two spinor waves are superposed are described with particular reference to neutron interferometry, and we examine what happens during an experiment designed to measure the spin observable, a process that is usually discussed informally in terms of projecting a system into an eigenstate of the measured observable. We shall see that the outcome of such measurements, apparently irreducibly discrete processes, can be described in terms of a continuous correlated evolution of the position and spin of a particle. Finally, we apply our theory to the Einstein-Podolsky-Rosen (EPR) spin experiment and show how correlated evolutions in the properties of distantly separated systems come about by the action of the non-local quantum potential and quantum torque. We do not discuss here developments in the relativistic theory¹⁹⁻²¹.

Causal theory of spin- $\frac{1}{2}$ particles

In the case of a spin-zero Schrödinger particle it is useful to interpret the physically real field $\Psi(\mathbf{x}, t)$ in terms of the coupled amplitude and phase functions $R(\mathbf{x}, t)$ and $S(\mathbf{x}, t)$, where $\Psi = R e^{iS/\hbar}$. The particle trajectories may be found by solving the equation

$$\frac{d\mathbf{x}}{dt} = (1/m) \nabla S|_{\mathbf{x}=\mathbf{x}(t)}$$

for $\mathbf{x}(t)$, once the initial position $\mathbf{x}(0)$ is specified (m is the mass of the particle). In the causal interpretation of the Pauli equation one aims at generalizing this procedure to interpret all four real

components in the spinor wavefunction $\Psi^a(\mathbf{x}, t)$ ($a = 1, 2$) in a visualizable way by constructing a mathematically equivalent set of fields in euclidean space. There are various ways of doing this. One way is to write

$$\Psi^a = R e^{iS/\hbar} \phi^a \quad (1)$$

where $S/\hbar$ is the average phase of Ψ and ϕ is a two-component object satisfying the conditions that $\phi^\dagger \phi = 1$ and that its average phase is zero. The two real degrees of freedom in ϕ may be interpreted through the properties of a spin vector field

$$\mathbf{S} = \frac{\hbar}{2} \phi^\dagger \boldsymbol{\sigma} \phi, |\mathbf{S}| = \hbar/2 \quad (2)$$

where $\boldsymbol{\sigma}$ represents the three Pauli matrices. Equation (1) is not a covariant decomposition because S is not a scalar quantity under spinor transformations, but it is valuable for understanding the content of the wave equation. In fact, S is essentially an angle, as follows from the well known representation of a spinor in terms of the Euler angles (θ, ψ, χ) (ref. 22)

$$\Psi = R e^{i\chi/2} \begin{pmatrix} \cos \theta/2 e^{i\psi/2} \\ i \sin \theta/2 e^{-i\psi/2} \end{pmatrix} \quad (3)$$

In this representation the spin vector (2) takes the form

$$\mathbf{S} = \frac{\hbar}{2} (\sin \theta \sin \psi, \sin \theta \cos \psi, \cos \theta). \quad (4)$$

The physical model we shall use to describe a spin- $\frac{1}{2}$ particle is as follows. We suppose that the spinor wavefunction defines a physically real multi-component field whose structure may be understood in terms of the interconnected functions $\rho (= R^2)$, S and $\mathbf{S}$, or ρ , θ , ψ and χ . In addition to this wave there is a small rigid spinning body which moves within the wave and is guided by it. The centre of mass of the body pursues a track $\mathbf{x} = \mathbf{x}(t)$ and we suppose that its internal angular momentum vector is just the spin vector field $\mathbf{S}$ evaluated along the trajectory. As well as defining properties (velocity and spin) associated with the corpuscle, the field variables enter into potential, force and torque terms in the equations of motion and precession of the spinning body. Moreover, these variables have other properties; the function ρ is the probability density of an ensemble of particles associated with the same wavefunction and S has the role of a Hamilton-Jacobi action function for the translational motion of the particle.

To see this, we substitute equation (3) into the Pauli equation

$$i\hbar \frac{\partial \Psi^a}{\partial t} = \left\{ \left(-\frac{\hbar^2}{2m} \left(\nabla - \frac{ie}{\hbar c} \mathbf{A} \right)^2 + eA_0 + V \right) \delta_b^a + \mu B_i \sigma_i^a{}_b \right\} \Psi^b \quad (5)$$

where $\mathbf{A}$ and A_0 are the external electromagnetic potentials, $\mathbf{B} = \nabla \times \mathbf{A}$ and V is any other external potential. One may write this in a closed form in terms of our chosen variables (or in terms of still more representations of Ψ).

We find a continuity equation for the density ρ .

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (6)$$

where

$$\mathbf{v} = \frac{\hbar}{2m} (\nabla \chi + \cos \theta \nabla \phi) - \frac{e}{mc} \mathbf{A} \quad (7)$$

is the particle velocity, and a generalized Hamilton-Jacobi equation

$$\frac{\hbar}{2} \left(\frac{\partial \chi}{\partial t} + \cos \theta \frac{\partial \phi}{\partial t} \right) + \frac{1}{2} m v^2 + Q + Q_s + \frac{2\mu}{\hbar} \mathbf{B} \cdot \mathbf{S} + eA_0 + V = 0. \quad (8)$$

Here, $(\hbar/2)\chi$ is identified with the action function,

$$Q = -\frac{\hbar^2}{2m} \frac{\nabla^2 R}{R} \quad (9)$$

is the usual quantum potential and

$$Q_s = -\frac{\hbar^2}{8m} ((\nabla\theta)^2 + \sin^2\theta(\nabla\phi)^2) \quad (10)$$

is a spin dependent addition. The two remaining equations of motion, for the variables θ and ϕ may be written most conveniently in terms of the spin vector (3):

$$\frac{d\mathbf{s}}{dt} = \mathbf{T} + \frac{2\mu}{\hbar} \mathbf{B} \times \mathbf{S} \quad (11)$$

where

$$\mathbf{T} = \frac{1}{mp} \mathbf{S} \times \partial_i (\rho \partial_i \mathbf{S}) \quad (12)$$

is an additional quantum torque which introduces a quantum mechanical precession of the spin vector along a trajectory ($d/dt = \partial/\partial t + \mathbf{v} \cdot \nabla$).

Clearly, terms involving θ and ϕ in the equations above imply purely spin-dependent effects not present in the Schrödinger theory. From equation (7), we see that the velocity contains a spin-dependent contribution, as does the energy

$$E = \frac{\hbar}{2} \left(\frac{\partial\chi}{\partial t} + \cos\theta \frac{\partial\phi}{\partial t} \right). \quad (13)$$

We also see that where Q and Q_s can be neglected in comparison with other energies, we obtain from equation (8) the Hamilton-Jacobi equation for a classical spinning particle. This establishes the classical limit of the theory: it is not necessary to assume that $\hbar \rightarrow 0$.

Once the initial spinor wavefunction is specified, we can find the translational motion $\mathbf{x} = \mathbf{x}(t)$ of the particle by solving

$$\frac{d\mathbf{x}}{dt} = \mathbf{V}(\mathbf{x}, t) \Big|_{\mathbf{x}=\mathbf{x}(t)} \quad (14)$$

given the initial position $\mathbf{x}_0 = \mathbf{x}(t_0)$. The trajectories will in general differ from those implied by the Schrödinger equation, even when there are no external potentials or when $e = \mu = 0$ (so that each spinor component in equation (5) evolves independently) for in this case the equation of motion derived from equation (8) is

$$m \frac{dv_i}{dt} = -\partial_i Q - \frac{1}{mp} \partial_k (\rho \partial_i S_j \partial_k S_j). \quad (15)$$

A spin-dependent force appears in addition to the usual quantum force derived from Q . Similarly, in the free case the spin vector will precess as a result of the torque, according to equation (12). This occurs when the wavefunction is a superposition of eigenstates of the spin operator $(\hbar/2)\boldsymbol{\sigma} \cdot \mathbf{n}$ in some direction $\mathbf{n}$ so that the spin degrees of freedom are functions of $\mathbf{x}$ and t . This is the analogue of the fact that the trajectory of a free particle in the Schrödinger case is not rectilinear if the wavefunction is a superposition of states of different momenta (a wave packet).

For the theory based on the functions ρ , θ , ϕ and χ to be equivalent to that based on ψ , we have to supplement equations (6), (8) and (11) with the following conditions on the Euler angles:

$$\oint d\chi = 2\pi n_1, \quad \oint d\theta = 2\pi n_2, \quad \oint d\phi = 2\pi n_3$$

where n_i are integers.

The quantum potential (9) and quantum torque (12) are highly non-classical in nature, in that they depend on all the parameters associated with a given process, including the experimental conditions and those associated with the particle whose motion they guide. Although they account for all the quantum features of spin they do not therefore have a universal form (in contrast to external classical potentials whose form is independent of, and external to, the context).

The above theory may be generalized to a many-body system¹⁷. Here we shall consider a system of two spin- $\frac{1}{2}$ particles. The wavefunction is now defined on configuration space and carries two spin indices: $\psi^{ab} = \psi^{ab}(\mathbf{x}_1, \mathbf{x}_2, t)$. It thus has eight real components. The configuration space dependence implies that we cannot interpret the wavefunction as a wave propagating in three-dimensional euclidean space as was possible in the one-body case, but we can nevertheless retain the notion of (correlated) individual well defined particle trajectories in three-dimensional space and endow these moving particles with further properties such as spin angular momentum. The irreducible necessity of using waves in multi-dimensional spaces to account for the correlated motions of the particles is one of the most striking features of quantum mechanics. It is the proposal of the causal interpretation that we should ascribe as much physical reality to such waves as we do to ones in euclidean space, and that they manifest themselves indirectly through the quantum potential and quantum torque by the influence they exert on the real motions and real physical properties of correlated particles moving in three-dimensional space.

To give a meaning to the two-body wavefunction, we proceed as above and construct an equivalent set of functions which are more easily visualizable. These are of course defined on configuration space as well. First of all, as in equation (1) we write

$$\psi^{ab} = R e^{iS/\hbar} \phi^{ab} \quad (16)$$

where R and S are real amplitude and average phase functions respectively, and ϕ satisfies $\phi^\dagger \phi = 1$ and has zero average phase. By $\phi^\dagger$, we mean here ϕ^{*ab} . To interpret the six degrees of freedom in ϕ^{ab} , we construct two spin vectors analogous to equation (2):

$$S_{1j} = \frac{\hbar}{2} \phi^\dagger e_{1j} \phi, \quad S_{2i} = \frac{\hbar}{2} \phi^\dagger e_{2i} \phi \quad (17)$$

where $e_{1i} = \sigma_i \times I$ and $e_{2i} = I \times \sigma_i$ are two sets of generators of two commuting Pauli algebras, and I is the unit 2×2 matrix. Taken together these vectors comprise only five independent degrees of freedom, but as

$$|S_1| = |S_2| \quad (18)$$

The remaining degree of freedom may be described with the aid of a local spin correlation tensor

$$S_{ij} = \frac{\hbar}{2} \phi^\dagger e_{1j} e_{2i} \phi \quad (19)$$

The 16 real tensor quantities $\rho (= R^2)$, S_{1i} , S_{2i} and S_{ij} are interconnected by a set of identities so that only seven of them are independent¹⁷. Together with S they cover all the degrees of freedom in the wavefunction and help to explain its meaning. If a completely covariant description is required we may interpret S in terms of the orientation of a further set of vectors constructed from ψ , but we shall not do so here.

For clarity we shall suppose that the particles have magnetic moments μ_1 and μ_2 but are electrically neutral and only subject to magnetic fields. The two-body Pauli equation is:

$$i\hbar \frac{\partial \psi^{ab}}{\partial t} = \left\{ \left(-\frac{\hbar^2}{2m_1} \nabla_1^2 - \frac{\hbar^2}{2m_2} \nabla_2^2 \right) \delta_c^a \delta_d^b + \mu_1 B_{1i} e_{1i cd}^{ab} + \mu_2 B_{2i} e_{2i cd}^{ab} \right\} \psi^{cd} \quad (20)$$

This may be written in a closed form in terms of the above variables. As in the one-body case, in addition to defining physically effective fields which enter into the equations of motion and precession of the particles as causal agents, these functions have other interpretations. Thus, ρ will be the probability density of an ensemble of systems. It should be noted that the factor ρ in ψ is not *a priori* a probability distribution but a real function in configuration space which determines the quantum potential. One may justify its interpretation as a probability density (of correlated particle positions) by the introduction of

subquantal fluctuations, because ρ satisfies the conservation equation (21). Indeed if one assumes that equation (21) is preserved by these fluctuations, one sees (following exactly in configuration space the steps of Bohm and Vigier's demonstration in 3-space²³) that the fluctuations force any arbitrary distribution into a distribution described by ρ . In addition, S is a Hamilton-Jacobi function for the translational motion. We have

$$\frac{\partial \rho}{\partial t} + \nabla_1 \cdot (\rho \mathbf{v}_1) + \nabla_2 \cdot (\rho \mathbf{v}_2) = 0 \quad (21)$$

$$\begin{aligned} \frac{\partial S}{\partial i} - i\hbar \phi \frac{\partial \phi}{\partial t} + \frac{1}{2} m_1 v_1^2 + \frac{1}{2} m_2 v_2^2 + Q_1 + Q_2 + Q_{1s} + Q_{2s} \\ + \frac{2\mu_1}{\hbar} \mathbf{B}_1 \cdot \mathbf{S}_1 + \frac{2\mu_2}{\hbar} \mathbf{B}_2 \cdot \mathbf{S}_2 = 0 \end{aligned} \quad (22)$$

where

$$\mathbf{v}_\mu = \frac{1}{m_\mu} (\nabla_\mu S - i\hbar \phi^\dagger \nabla_\mu \phi), \mu = 1, 2 \quad (23)$$

are the velocities of the particles, $Q_\mu = (-\hbar^2/2m_\mu) \nabla_\mu^2 R/R$ are the usual quantum potential that arise in the spin-zero two-body problem, and

$$Q_{\mu s} = \frac{1}{4m_\mu} [\partial_{\mu i} S_{jk} \partial_{\mu i} S_{jk} + \partial_{\mu i} S_{1j} \partial_{\mu i} S_{1j} + \partial_{\mu i} S_{2j} \partial_{\mu i} S_{2j}] \quad (i = 1, 2)$$

are spin-dependent additions. The remaining information in equation (20) may be expressed in terms of the evolution of S_{1i} , S_{2i} and S_{ij} . These are somewhat complicated and we shall only give the equations for the spin vectors

$$\frac{d\mathbf{S}_\nu}{dt} = \mathbf{T}_\nu + \frac{2\mu_\nu}{\hbar} \mathbf{B}_\nu \times \mathbf{S}_\nu, \nu = 1, 2 \quad (24)$$

where $\mathbf{T}_\nu$, $\nu = 1, 2$, are quantum torques given by

$$\begin{aligned} T_{1\kappa} = \frac{1}{2\rho m_1} \varepsilon_{ijk} \{S_{1i} \partial_{1l} (\rho \partial_{1l} S_{1j}) + S_{1r} \partial_{1l} (\rho \partial_{1l} S_{jr})\} \\ + \frac{1}{2\mu m_2} \varepsilon_{ijk} \{S_{1i} \partial_{2l} (\rho \partial_{2l} S_{1j}) + S_{1r} \partial_{2l} (\rho \partial_{2l} S_{jr})\} \quad (25) \\ T_{2\kappa} = [1 \leftrightarrow 2] \end{aligned}$$

and $d/dt = \partial t + \mathbf{v}_1 \cdot \nabla_1 + \mathbf{v}_2 \cdot \nabla_2$.

The properties of analogous entities in the one- and two-body theories are quite different. For example, the magnitudes of the spin vectors in equation (17) are continuously variable and indeed may be zero in regions where $\rho \neq 0$. In contrast the spin vector in equation (2) has a fixed magnitude. The reason for this difference is that all the functions of interest are constructed out of the total quantum state associated with a given system, and not arbitrarily specified.

Given a solution to the wave equation we simply substitute into equation (17) to find the spin vector components, and into equation (23) to find the velocities. These functions are in general dependent on both sets of coordinates. The particle trajectories, defined by the solutions to

$$\mathbf{v}_\mu = \frac{d\mathbf{x}_\mu}{dt}, \mu = 1, 2 \quad (26)$$

(given both initial positions) and the spin vector orientations, therefore depend irreducibly on each other and on the total quantum state. The trajectories and spin vectors will only evolve independently when the wavefunction factorizes, as

$$\psi_{ab}(\mathbf{x}_1, \mathbf{x}_2, t) = \psi_{1a}(\mathbf{x}_1, t) \psi_{2b}(\mathbf{x}_2, t). \quad (27)$$

These correlations in the two motions are brought about by the spin-dependent quantum potentials and quantum torques, that are functions of both coordinates. They decompose into a sum of functions each associated with just one of the particles when

equation (27) holds. We shall illustrate this action, which under certain conditions can extend to indefinitely large distances, under circumstances where one can be certain that the individual particle packets do not overlap. Such a possibility of non-local action may be seen for example from the intertwined character of the equation of precession (24).

It is emphasized that in the approach proposed here, it is possible to distinguish the particles in a many-body system by their individual trajectories (including when the particles are identical in the quantum mechanical sense).

Finally, it should be noted that the causal interpretation described in this section is appropriate to just the theory of spin in its Pauli form, and other approaches are possible if we work in a different representation of the underlying quantum theory^{17,18}. For example, one can arrive at a more detailed theory if a further set of parameters are introduced to describe the intrinsic orientation of a body, in the same way as the coordinates $\mathbf{x}$ may be associated with its intrinsic position. The present theory results when one averages over these new variables. Strictly speaking, this more complete theory is necessary in the many-body case in order that we may speak of spin vectors intrinsically connected with each of the bodies, but our approach in this paper nevertheless gives an initial insight into spin-dependent phenomena.

Spin superposition in neutron interferometry

It is a well-known mathematical result that the resultant spin vector of a superposition of spin-up and spin-down eigenstates of the spin operator in some direction lies in a plane orthogonal to that of the spin vectors associated with the constituent spinors. This follows easily from the definition (2). This prediction of the linear superposition law for spinors has been experimentally tested and verified in single-crystal neutron interferometry²⁴⁻²⁶. We shall describe here a simple model which shows how the results of this experiment may be understood through the action of the quantum torque (12).

A neutron beam polarized in the z -direction enters the interferometer and is split into an upper (2) and lower (1) component. The upper beam is passed through a d.c. spin-flipping coil whose homogeneous magnetic field couples with the magnetic moment μ and rotates the spin vector about the y -axis (in a right handed coordinate system). The magnetic field and period of interaction are fixed so that the spin vector reverses. The two beams, with opposite spin and a relative phase difference λ introduced by a scalar phase-shifter, converge coherently on the final set of crystal planes, where each beam is split once more. As a result, the forward and deviated beams emerging from the interferometer both contain a coherent superposition of components of opposite spin. It is this final stage of the process that we shall describe.

The final crystal planes may be treated as a semi-transparent mirror, represented by a square potential barrier V (ref. 9). The initial two-component spinor (that is, just before the final semi-transparent surface) is a superposition of two wavepackets associated with the spin-up and spin-down beams:

$$\psi_0 = f_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + f_2 e^{i\lambda} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (28)$$

where f_1 and f_2 are Gaussian packets. The spin vectors (2) associated with each beam are then

$$\mathbf{S}_1 = \frac{\hbar}{2} (0, 0, 1), \quad \mathbf{S}_2 = \frac{\hbar}{2} (0, 0, -1). \quad (29)$$

Figure 1 shows the trajectories that result from solving equation (14) numerically with a range of initial positions (remember that only one of these is realized in any given experiment). The horizontal lines indicate the position of the square potential and the coordinate y is parameterized by time t . Notice that all the trajectories originating in the upper (lower) beam inside the

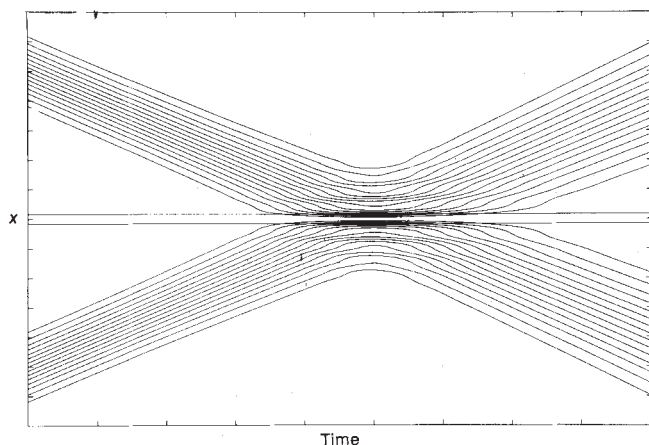


Fig. 1 Particle trajectories at the last semi-transparent surface of a neutron interferometer with one spin-rotating coil. Units are arbitrary.

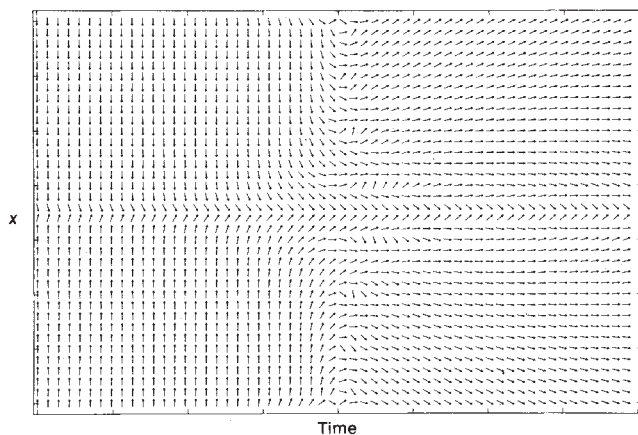


Fig. 2 The polar orientation angle of the spin vector $\theta(x, t)$ plotted on the mesh of points of Fig. 1.

interferometer enter the upper (lower) beam outside. The polar orientation field $\theta(x, t)$ is shown in Fig. 2, from which it is clear that the resultant spin vector in the emerging beams lies in the xy -plane ($\theta = \pi/2$). Superposing the trajectories mentally on this figure allows us to see the way in which the spin orientation changes in a continuous manner during the scattering from the potential through the action of the quantum torque. Changing the relative phase λ does not affect the polar angle, but causes the final direction of the spin vector in the xy -plane to be rotated. Notice that once the spin vector has been rotated into the xy -plane in each beam, it stays there, because each emerging beam is composed of transmitted and reflected parts of the two incoming packets.

The spin rotation on beam 2 may also be carried out with a time-dependent radio frequency spin coil²⁷. This produces a time-dependent relative phase $\lambda\omega_{rf}t$ so that the final direction of the spin vector in the xy -plane depends on the phase of the coil at the time the wavepacket passed. Of course the time dependence only shows up in an ensemble of cases; each individual neutron has a static spin orientation outside the interferometer. In the case where a coil is placed in each of the beams^{28,29}, one obtains just spatial interference in the final crystal planes and the spin vector remains constant (in the z -direction).

The final stage of these experiments, the measurement of the spin, is discussed in the next section.

Spin measurement

In the causal interpretation, matter possesses well defined attributes regardless of whether or not an observation is being made. As a result, the theory has a less obsessive concern with the act of measurement, and treats this as a particular case of the general physical processes that it describes. In fact, it shows why the word 'measurement' is a misleading one in a quantum context since 'measuring an observable' is not an act which simply reveals an aspect of an independent reality, with the measuring instrument playing no essential role in the determination of the outcome of an experiment as in classical physics. Rather, the observing system and the object of observation, which have an equal ontological status, are to be thought of as two systems with continuous trajectories in spacetime, that interact for a while, in the process mutually transforming one another, and then cease to interact. The outcome of the 'measurement' is thus a transformation of both irreducibly coupled systems; the action of the apparatus cannot be separated out and 'calcu-

lated away'. The causal interpretation therefore clarifies what Bohr meant by the indivisible wholeness of observer and observed.

To illustrate this, let us consider what is involved in a 'measurement of the spin observable' in a Stern-Gerlach experiment^{30,31}. Roughly speaking, a typical account of this process proceeds as follows. When the spin along a given direction, say z , is measured, by passing a beam through a Stern-Gerlach device, the wavefunction is 'projected' into an eigenstate of the spin operator $(\hbar/2)\sigma_z$ in that direction. That is, the spin becomes definite in the z -direction, with two possible values $\pm\hbar/2$, whereas an uncertainty is introduced in the two mutually perpendicular directions (the spin components in these two directions are to be thought of as 'randomly fluctuating'). Thus, before measurement the spin is a 'potentiality' and yet even after measurement, the spin vector cannot be considered to be definite. As regards the result that is actually realized in any given experiment, the most that one can do is to state the probabilities of spin-up and spin-down.

There is an evident lack of clarity as to what all this means physically in each individual experiment and in particular how the apparent discreteness (up or down, nothing in between) comes about. To see how a causal description may be given, suppose that a polarized beam of spin- $\frac{1}{2}$ particles (with an intensity so low that only one particle is present at any one time) passes between the poles of a magnet which has a gradient in the z -direction. To achieve an unambiguous outcome of the experiment, we suppose that the initial wavefunction ψ_0 has the profile of a gaussian packet $f(z)$ in the z -direction and that the spin- $\frac{1}{2}$ particle has a definite position within it. We are only interested in the motion along the direction of the analysing field and so shall suppress the y -motion (which is parameterized by t). We have

$$\psi_0 = f(z) \begin{pmatrix} c_+ \\ c_- \end{pmatrix} \quad (30)$$

where c_+ , c_- are constant coefficients (which we take to be real) which fix the initial definite orientation of the spin vector:

$$\theta_0 = \cos^{-1}(c_+^2 - c_-^2), \quad \phi_0 = \frac{\pi}{2}, \quad \chi_0 = -\frac{\pi}{2} \quad (31)$$

(where we have compared equations (30) and (3)) and determine the probabilities of the possible outcomes, c_+^2 (spin-up) and c_-^2 (spin-down). The initial velocity in the z -direction is zero from equation (7).

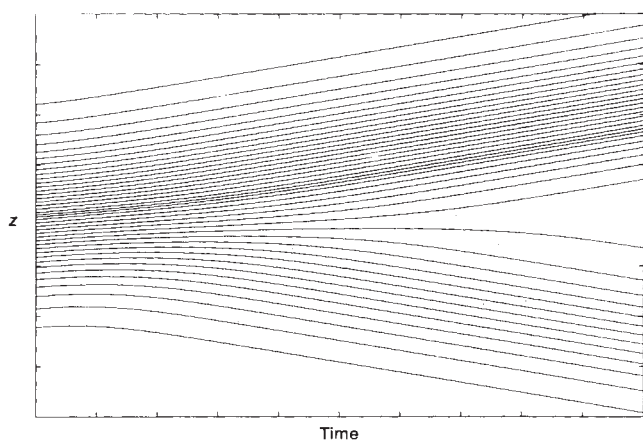


Fig. 3 The trajectories for a Gaussian distribution of initial positions in the region of wave packet separation during a spin measurement with $c_-^2 = 0.25$ and $c_+^2 = 0.75$.

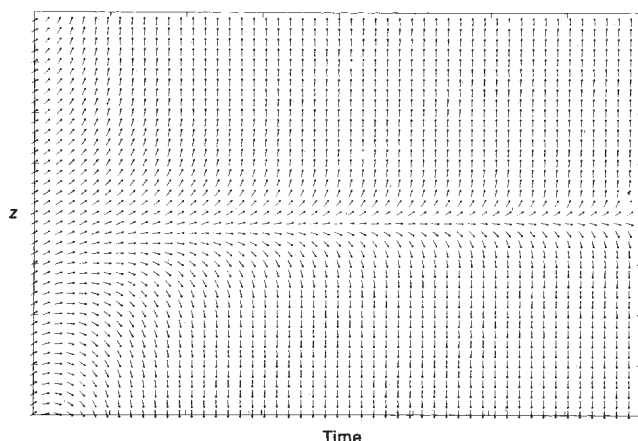


Fig. 4 The field of orientations $\theta(z, t)$ associated with Fig. 3.

The effect of the impulsive action of the magnetic field is to create two wave packets which begin to separate in the z -direction. At time t after leaving the magnet, the wavefunction found by solving the Pauli equation (5) is¹⁴

$$\Psi(z, t) = \begin{pmatrix} c_+ f_+(z, t) \\ c_- f_-(z, t) \end{pmatrix} \quad (32)$$

where $f_{\pm}(z, t)$ are packet functions. From this we can calculate ρ , θ , ϕ and χ for all z and t , and hence plot trajectories and spin vectors for various initial positions and orientations (given appropriate choices of the apparatus parameters). The particle joins one of the packets, and the experimental parameters are fixed so that the packets separate to a macroscopic distance, with negligible overlap, so that when the particle is incident on a detecting screen an unambiguous result is obtained. Observing the position of the particle (the spot on the screen) allows us to infer the value of the spin, and so the 'observed system' is the spin variable, and the coordinates of the 'observing apparatus' comprise the particle position. Which packet the particle actually enters is determined solely by its initial position before entering the magnet, and the initial spin vector orientation.

When $c_+ = 0$ ($c_- = 0$) all the trajectories move upward (downward) to reach the upper (lower) portion of the detecting screen, and the polar angle maintains a constant value of $\theta = 0(\pi)$. The initial state is an eigenstate of σ_z and the particle enters the upper (lower) path regardless of its initial position.

Figure 3 shows a selection of trajectories beyond the exit to the magnet for the case $c_+^2 = 0.75$, $c_-^2 = 0.25$, when the spin vector initially makes an angle $\theta_0 = 60^\circ$ with the z -axis. This initial condition fixes the position of the bifurcation point of the two packets on the z -axis; particles initially above (below) this point join the upward (downward) moving packet. It will be observed how the usual probabilities of the outcomes (up or down) are reflected in the relative proportions of trajectories entering the two packets (recall that only one trajectory is realized in any individual experiment, and the statistical nature of the results arises from the practical uncontrollability of the initial position).

Figure 4 depicts the polar angle $\theta(z, t)$ of the spin vector. As the packets begin to separate and the trajectories are bunched into two channels, the quantum torque acts to rotate the spin vector so that eventually it is pointing in the $\pm z$ -direction (and never in any other direction). We therefore have a model that yields the usual discrete results of quantum mechanics, but in a way that allows us to understand these results as the outcome of an objective continuous evolution of the properties of matter

which exist independently of the act of observation. The final irreversible stage of the measurement process, the detection by a screen, merely tells us what has already happened and there is no need to ascribe to this act of detection a particular significance in the sense that only when the particle is detected can its spin in the relevant direction be said to be definite (the so-called 'reduction of the wavepacket').

From the above we see that it is possible to retain the notion of definite particle trajectories and well-defined spin vectors throughout the process because the spin-dependent quantum potential and quantum torque account perfectly for the quantum features of spin. Whereas classically the attempt to align a spin vector with an inhomogeneous magnetic field simply results in a torque $\propto (\mathbf{S} \times \mathbf{B})$, which makes the spin vector precess about the field, the quantum torque produces an extra rotation which actually aligns the spin vector in one direction or the other ($\theta = 0$ or π). Whether the spin vector associated with a particular trajectory becomes aligned with or opposed to the field depends on the initial orientation with respect to the field and the initial position in the wave packet.

Notice that changing the orientation of the Stern-Gerlach magnet relative to the direction of polarization of the incoming beam (which amounts to measuring a different component of the spin observable) gives rise to different spin dependent quantum potentials and consequently different correlated evolutions of the trajectories and spin vectors. It is in this way that the causal interpretation gives an objective basis to the uncertainty relations. Two settings of the magnet implies two mutually exclusive apparatuses and therefore different total wavefunctions. In a statistical ensemble of experiments the usual Heisenberg correlation between the dispersions of the spin observables in the two directions will be obtained, but this has nothing to do with our ability to define precisely all components of the spin vector associated with a particle before, during and after an experiment.

Experiment of Einstein, Podolsky and Rosen

In 1935, Einstein, Podolsky and Rosen (EPR) published an article which has had a deep influence on the debate on the interpretation of quantum mechanics. The argument of EPR was originally directed at showing the incompleteness of quantum mechanics, but it has increasingly been viewed as demonstrating the non-local nature of certain quantum many-body systems (that is, that the properties of distantly separated systems not in spatial contact and not classically interacting can depend on one another). We shall discuss this experiment in the version devised by Bohm³⁰. This is theoretically very similar to the

photon polarization experiments that have been carried out to test the predictions of quantum mechanics in this case^{33,34}.

A pair of spin- $\frac{1}{2}$ particles of mass m and magnetic moment μ are formed in a confined region in a simultaneous eigenstate of the spin operator in the z -direction $(\hbar/2)(\sigma_{z_1} + \sigma_{z_2})$ and the total spin operator $[\hbar/2(\boldsymbol{\sigma}_1 + \boldsymbol{\sigma}_2)]^2$ with eigenvalue zero. The particles separate in the y -direction and pass through gaussian slits oriented to produce packets in the directions of the analysing fields of two identical Stern–Gerlach devices. The magnet 2 is set to measure the spin in the z -direction, and the magnet 1 has been rotated counterclockwise through an angle δ about the y -axis so that it has a gradient in the z' -direction.

At the entrance to the magnets the wavefunction is

$$\psi_0 = f_1(z'_1)f_2(z_2) \frac{1}{\sqrt{2}}(u_+v_- - u_-v_+) \quad (33)$$

where $f_1(z'_1)$ and $f_2(z_2)$ are Gaussian packets, z'_1 and z_2 are the coordinates of particles 1 and 2 in the z' and z directions respectively, and $\sigma_{z_1}u_{\pm} = \pm u_{\pm}$, $\sigma_{z_2}v_{\pm} = \pm v_{\pm}$. The state (33) predicts the following expectation value for the correlations of spins measured in the z and z' directions:

$$\langle \sigma_{z_1}\sigma_{z_2} \rangle = -\cos \delta, \quad (34)$$

and the probabilities of the possible outcomes are given by

$$P(++) = P(--)=\frac{1}{2}\sin^2\frac{\delta}{2}, \quad P(+-)=P(-+)=\frac{1}{2}\cos^2\frac{\delta}{2} \quad (35)$$

In equation (33) we have suppressed the motion in the y -direction because this is not relevant to the measurement process. We assume only that the particles are sufficiently far apart on the y -axis so that they do not interact in the usual sense, and so that the measuring devices cannot influence one another.

As with the case described in the previous section, equation (33) describes a state in which the spin is independent of position and the purpose of the Stern–Gerlach devices is to introduce couplings between the spins (the measured variables) and the positions (the apparatus coordinates). At the exit to each magnet two superposed packets are formed which separate in time along the directions of the analysing fields.

Before we describe how the correlations in the results of the two spin measurements are to be understood in our approach, let us first recall why the features brought to light by EPR are conventionally felt to be problematical, indeed paradoxical³⁵.

It is a straightforward prediction of the wavefunction (24) that if any component of the spin of particle 1 is measured and the result found to be $\pm(\hbar/2)$, then it can be immediately concluded that the same component of the spin of particle 2 (which is not measured) becomes definite and is equal to $\mp(\hbar/2)$. This will be true whatever component of particle 1 is measured since the singlet state is rotationally symmetric:

$$\frac{1}{\sqrt{2}}(u_+v_- - u_-v_+) = \frac{1}{\sqrt{2}}(u'_+v'_- - u'_-v'_+)$$

where $u'_i = a_{ij}u_j$ ($i, j = \pm$) and a_{ij} is a spinor rotation matrix.

Unlike classical physics, it is claimed that only one component of the quantum mechanical spin of each particle can have a definite value at a given time. The two perpendicular components are indeterminate and are to be thought of as 'randomly fluctuating'. But we are free to measure the spin of particle 1 in any direction we choose, and so make definite the spin of particle 2 in any direction. Moreover, this choice may be made during the flight of the particles. The paradox then arises in the circumstance that particle 2 must somehow 'know' in which direction it is to be definite and in which it is to be fluctuating.

EPR (ref. 32) used such an argument to show that quantum mechanics is incomplete; there should be an 'element of reality' corresponding to the spin of each particle since apparently

without in any way disturbing one of the particles we can predict with certainty the value of its spin in any direction (having performed a measurement on the other particle). There is however no way in the usual formalism based on wavefunctions and operators to ascribe a simultaneous reality to spin components in perpendicular directions.

Bohr's answer³⁶ was that the form of the experiment and the content of the results form an unanalysable whole and that no paradox will arise if we refrain from drawing inferences from the results. The choice of different analysing field directions implies a series of mutually exclusive experimental arrangements whose details cannot be compared with one another.

The problem may be resolved in a different manner, however, in the causal interpretation which, as we have seen, adopts Einstein's view that there is a reality independent of measurement and shows that this idea is completely consistent with the quantum mechanical formalism. The error in the conventional formulation of the 'paradox' outlined above is to suppose that if a reality is to be ascribed to entities such as 'spin', then this reality refers to the eigenvalues of operators; that is to suppose that the dynamical variable is only realized and becomes definite (equal to an eigenvalue) on performing a measurement.

The velocities and spin vectors may be calculated from equations (17), (23) and (33) to be $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{S}_1 = \mathbf{S}_2 = 0$. It may appear strange that the particles have zero true values of internal angular momentum but it should be borne in mind that in quantum mechanics the properties of individuals depend on the state of the whole of which they are a part. Each individual spin vector will possess properties different to those of the spin vector associated with a single body (which by equation (2) can never be zero). Notice in particular that the spins of the particles in each separating packet are not determined by either of the addends in the state (33)—they are not in an initial state in which the spin of one is up (down) and the other is down (up), as one might expect in the analogous classical case.

Immediately after the interaction with the magnets, however, the particles acquire non-zero z' and z components of their velocities and the quantum torques $\mathbf{T}_1$ and $\mathbf{T}_2$ (equation 25) act to start the particles spinning. In general, the particle coordinates and spin vectors have a complicated dependence on one another¹⁶ and which packet each particle enters at the exits to the magnets, and hence which result $\pm(\hbar/2)$ is obtained for S_{z_1} and S_{z_2} , depends sensitively on the initial positions of both. On the average the correlations yield expression (34) and the probabilities (35) are reflected by the relative proportions of trajectories that go up or down.

To gain insight into this non-local action, we shall consider two special cases for which we explicitly plot trajectories and spin vectors. First, let us suppose that a measurement is made only on one side. The calculation yields the result that at the exit of the field, the particle 1 undergoing measurement enters one of the separating packets, depending on its initial position on the z' -axis, under the influence of the total spin-dependent quantum potential, and its spin vector component in the direction of the analysing field changes continuously from 0 to $-\hbar/2$ or $+\hbar/2$ as the packets become separated in the z' -direction. Simultaneously the spin vector component in the z' -direction of the second particle, not undergoing measurement, changes in the opposite sense from 0 to $+\hbar/2$ or $-\hbar/2$ as a result of the operation of the non-local quantum torque, regardless of its position. That is, the spin of particle 2 depends on the position and hence spin of particle 1. The velocities however in this case remain independent; the trajectory of particle 2 is not affected by the measurement on particle 1 and the trajectory of particle 1 depends simply on local factors. If we were subsequently to measure the spin of particle 2 in the z' -direction, we would, of course, find the opposite result to that found for particle 1. The trajectories and spin vector magnitudes (indicated by the length of the arrow which always lies in the z' -direction) are shown in Fig. 5.

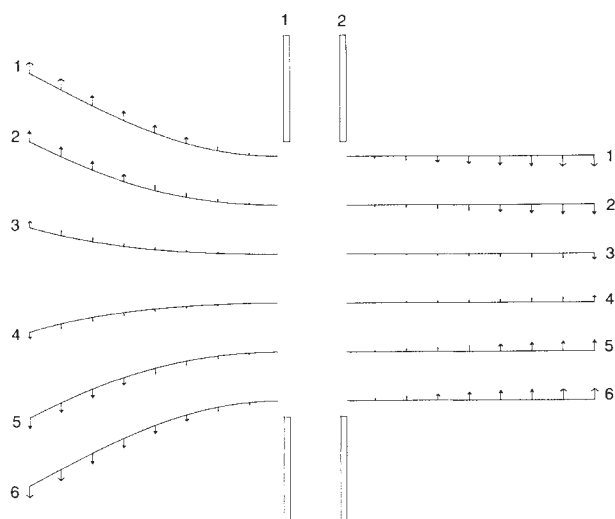


Fig. 5 Trajectories and correlated spin vector orientations for two particles initially in a singlet state after the impulsive measurement of the z' component of the spin of particle 1 only. The vertical lines labelled 1 and 2 represent Stern-Gerlach magnets. The size of the arrows indicates the magnitude of the spin.

Consider now the case in which both Stern-Gerlach devices are operational and set to measure the spin component in the z -direction ($\delta = 0$) on both particles simultaneously. The calculation in this case yields the results plotted in Fig. 6. These results are obtained by taking the initial position of particle 1 to be the same in each case and then calculating the correlated trajectories which develop for a representative range of initial positions of particle 2. The situation when the initial position of particle 1, $z_1(0)$, is chosen to be equal to that of particle 2, $z_2(0)$, represents a bifurcation point. If $z_2(0) < z_1(0)$ then particle 2 has a negative velocity and its z -spin component decreases from 0 to $-\hbar/2$ whilst the correlated particle 1 has a positive velocity and its z -spin component increases from 0 to $+\hbar/2$. Analogous correlations are found if $z_2(0) > z_1(0)$. Clearly the fate of each particle depends sensitively on what happens to the other one.

Returning to the general case ($\delta \neq 0$) it may be shown that whether one staggers the measurements in time, or performs them simultaneously, the statistical correlations will come out to be the same. The fate of a particular particle, however, will depend on when the measurements are carried out. Consider for example the correlated trajectories 3 in the figures. In Fig. 5, particle 1, will always move downward regardless of the initial position of particle 2. In Fig. 6, on the other hand, particle 2 moves up or down, depending on the value of $z_1(0)$. The identical statistics obtained in simultaneous or staggered measurements result in part, therefore, from quite different evolutions at the level of the individual particles which make up an ensemble. The non-local mechanism at work in the two cases is not the same, but they cannot be experimentally distinguished.

Let us summarize our model. We suppose that both particles always have well-defined trajectories and spin angular momenta components in all directions. The interaction of one of the particles with the Stern-Gerlach magnet implies the transformation of the wavefunction of the entire system, and consequently a change in the total spin-dependent quantum potential associated with the two bodies. As we have shown, when particle 1 interacts with the Stern-Gerlach device, it enters one or other of the emerging wave packets while the quantum torque T_1 rotates its spin vector so that it coincides with the eigenvalue of the spin operator being measured, and the quantum torque T_2 rotates the spin of particle 2 to an equal and opposite value while the trajectory is unaffected.

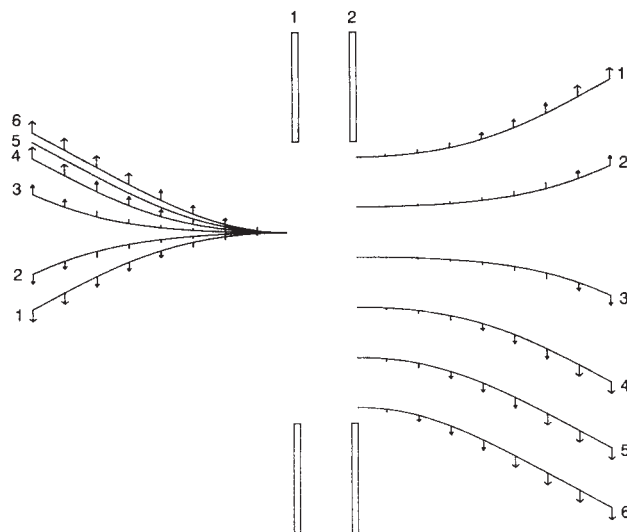


Fig. 6 Correlated pairs of trajectories and spin vector orientations after the impulsive measurement of the spin in the z direction on both particles $z_1(0)$ is constant, $z_2(0)$ is variable and the difference between them has been magnified for clarity. The magnets may in principle be separated by any distance.

The purpose of making simultaneous measurements on both particles is to show that these correlations in the particle's behaviour are brought about by an action-at-a-distance mechanism. Once the initial positions of the particles have been fixed, the outcome of the experiment is uniquely determined. The non-local spin-dependent quantum potential associated with the total quantum state rotates the spin vectors, as in the case of measurement on only one particle, and the total quantum potential ensures that the positions of the two particles are correlated in the required way. Changing the relative orientation of the magnets changes the total quantum state and hence implies different correlated motions. It is in this way that we incorporate the insights of Bohr concerning this process, but we also go on to explain how the correlations come about.

Although there is no classical interaction between the particles, they cannot be said to be non-interacting since they are tied by the quantum potential. Thus although we ascribe independent 'elements of reality' to individual quantum systems, the criterion for the existence of an independent 'element of reality' proposed by EPR (if without in any way disturbing a system we can predict with probability unity the value of a dynamical quantity, they that quantity is an element of reality) is not applicable in the causal interpretation. The act of measurement performed on one particle does indeed bring about a disturbance in the properties of the other particle, by means of the quantum potential.

Note that in bringing about these correlations, the quantum potential and quantum torque do not transfer information. We do not know what result would have been obtained for particle 2 had we not made a measurement on particle 1. There is therefore no conflict between this form of non-locality and the relativistic requirement that no signal be transmitted faster than the speed of light. Nevertheless, Einstein-Podolsky-Rosen-type correlations in the properties of distant systems may be felt to contradict at least the spirit of relativity which apparently requires actions only to be transmitted locally in order not to conflict with causality (that is, that effects follow causes). A detailed treatment of this problem using the techniques of modern predictive mechanics shows, however, that no such contradiction arises¹⁷. Remarkably, theories involving 'action-at-a-distance' may be compatible with the principles of relativity if they satisfy certain constraints.

Finally, our model throws light on the meaning of Bell's

inequality³⁷, a mathematical condition which must apparently be satisfied by any theory which seeks to account for quantum phenomena in terms of local 'hidden' parameters. We have discussed elsewhere the manner in which the causal interpretation violates Bell's condition¹⁶.

Discussion

The illustrations presented here demonstrate clearly how the results of experiments involving spin in quantum mechanics, including EPR-type experiments, can be accounted for in terms of a reality in which well-defined and continuously variable quantities evolve in a deterministic manner according to the equations of motion of the causal interpretation. They also illustrate that the fundamentally new feature of matter introduced in the quantum theory is a kind of wholeness in which the behaviour of an individual particle is irreducibly connected with its context (expressed through the wavefunction), evidenced in the two-particle case by the existence of non-local connections. Elements of reality exist but they are not described by the

quantum formalism on its own, which deals with eigenvalues of operators, and in general they are not simply revealed in measurement interactions. The interactions regarded as measurements are in fact those in which a particular variable of a 'measured' system becomes correlated with a particular apparatus coordinate according to deterministic laws of evolution of the whole undivided system-plus-apparatus. In this sense the elements of reality in quantum theory are essentially different to the elements of reality of newtonian physics. Although we use the same terms (for example, position, momentum, kinetic and potential energies) to describe the particle's motion, the individual and its relation with these attributes is not the same as that which exists in classical mechanics. In the history of science each new epoch-making discovery has indicated a new feature of matter, one of which was the introduction of the idea of the field. In our opinion, assuming reality has this fundamentally new feature of wholeness is preferable to assuming that it does not exist except when we are looking at it.

P.R.H. acknowledges the SERC for a European fellowship.

1. Jammer, M. *The Philosophy of Quantum Mechanics* (Wiley, New York, 1974).
2. Bohr, N. *Atomic Theory and the Description of Nature* (University Press, Cambridge, 1934); *Atomic Physics and Human Knowledge* (Wiley, New York, 1958).
3. de Broglie, L. *C. R. Acad. Sci. Paris* **183**, 447-448 (1926); **185**, 380-382 (1927).
4. de Broglie, L. *Non-Linear Wave Mechanics* (Elsevier, Amsterdam, 1960).
5. Bohm, D. *Phys. Rev.* **85**, 166-179, 180-193 (1952).
6. Philippidis, C., Dewdney, C. & Hiley, B. *Nuovo Cim.* **52B**, 15-28 (1979).
7. Dewdney, C. & Wiley, B. *Found. Phys.* **12**, 27-48 (1982).
8. Philippidis, C., Bohm, D. & Kaye, R. *Nuovo Cim.* **71B**, 75-84 (1982).
9. Dewdney, C. *Phys. Lett.* **A109**, 377-384 (1985).
10. Dewdney, C., Kyprianidis, A. & Vigier, J. P. *J. Phys. A*, **17**, L741-744 (1984).
11. Dewdney, C. & Holland, P. R. in *Quantum Mechanics versus Local Realism—The Einstein-Podolsky-Rosen Paradox* (ed. Selleri, F.) (Plenum, New York, in the press).
12. Bohm, D., Schiller, R. & Tiomno, J. *Nuovo Cim. Suppl.* **1**, 48-66 (1955).
13. Takabayasi, T. *Prog. Theor. Phys.* **14**, 283-302 (1955).
14. Dewdney, C., Holland, P. R. & Kyprianidis, A. *Phys. Lett.* **A119**, 259-267 (1986).
15. Dewdney, C., Holland, P. R. & Kyprianidis, A. *Phys. Lett.* **A121**, 105-110 (1987).
16. Dewdney, C., Holland, P. R. & Kyprianidis, A. *J. Phys. A*, **20**, 4717-4732 (1987).
17. Holland, P. R. *Phys. Rep.* (in the press).
18. Holland, P. R. *Phys. Lett.* **A128**, 9-18 (1988).
19. Holland, P. R. *Found. Phys.* **17**, 345-363 (1987).
20. Holland, P. R., Kyprianidis, A., Maric, Z. & Vigier, J. P. *Phys. Rev.* **A33**, 4380-4383 (1986).
21. Cufaro-Petroni, N., Dewdney, C., Holland, P. R., Kyprianidis, A. & Vigier, J. P. *Phys. Rev.* **D32**, 1375-1383 (1985); *Phys. Lett.* **A113**, 359-364 (1986).
22. Goldstein, H. *Classical Mechanics* 2nd edn (Addison-Wesley, Cambridge, 1980).
23. Bohm, D. & Vigier, J. P. *Phys. Rev.* **96**, 208-214 (1954).
24. Summhammer, J., Badurek, G., Rauch, H. & Kischko, J. *Phys. Lett.* **A90**, 110-112 (1982).
25. Badurek, G., Rauch, H., Summhammer, J., Kischko, U. & Zeilinger, A. *J. Phys. A*, **16**, 1133-1139 (1983).
26. Summhammer, J., Badurek, G., Rauch, H., Kischko, U. & Zeilinger, A. *Phys. Rev.* **A27**, 2523-2528 (1983).
27. Badurek, G., Rauch, H. & Summhammer, J. *Phys. Rev. Lett.* **51**, 1015-1018 (1983).
28. Dewdney, C., Garuccio, A., Kyprianidis, A. & Vigier, J. P. *Phys. Lett.* **A104**, 325-328 (1984).
29. Badurek, G., Rauch, H. & Tuppinger, D. *Phys. Rev.* **A34**, 2600-2608 (1986).
30. Bohm, D. *Quantum Theory* ch. 22 (Prentice-Hall, New York, 1951).
31. Böhm, A. *Quantum Mechanics* ch. 13 (Springer, Berlin, 1978).
32. Einstein, A., Podolsky, B. & Rosen, N. *Phys. Rev.* **47**, 777-780 (1935).
33. Aspect, A., Dalibard, J. & Roger, G. *Phys. Rev. Lett.* **49**, 91-94 (1982).
34. Perrie, W., Duncan, A. J., Beyer, H. J. & Kleinpoppen, H. *Phys. Rev. Lett.* **54**, 1790-1794 (1985).
35. Bohm, D. & Aharonov, Y. *Phys. Rev.* **108**, 1070-1076 (1957).
36. Bohr, N. *Phys. Rev.* **48**, 696-702 (1935).
37. Bell, J. S. *Physics* **1**, 195-200 (1964).

ARTICLES

A cloned octamer transcription factor stimulates transcription from lymphoid-specific promoters in non-B cells

Michael M. Müller, Siegfried Ruppert*, Walter Schaffner & Patrick Matthias

Institut für Molekularbiologie II der Universität Zürich, Hönggerberg, CH-8093 Zürich, Switzerland

* Deutsches Krebsforschungszentrum, im Neuenheimer Feld 280, D-6900 Heidelberg, FRG

The complementary DNA coding for a lymphocyte-specific transcription factor binding to the DNA 'octamer' sequence TNATTGTCAT has been cloned. The nucleotide sequence shows homology to the homoeobox domain. Expression of this cDNA in HeLa cells is sufficient for a strong transcriptional activation of B-cell-specific promoters.

THE 'octamer' motif TNATTGTCAT is found in all immunoglobulin promoters and also in the immunoglobulin heavy chain enhancer¹⁻³. B-cell-specific transcription of immunoglobulin genes has been shown to be critically dependent on the presence of an intact octamer motif⁴⁻⁶. At least three proteins that specifically recognize the octamer have been identified in nuclear extracts of mammalian cells: OTF-1 (an octamer-binding transcription factor, also referred to as NF-A1, ref. 7; OBP100, ref. 8; NF III, ref. 9) is a ubiquitous protein of apparent relative molecular mass, (M_r) 90,000-100,000 (90-

100 K)¹⁰. OTF-2A (or OTF-2, ref. 11; NF-A2, ref. 12) is restricted to lymphoid cells^{6,11-13} and has an apparent M_r of 60 K. OTF-2B is another lymphoid-specific protein of about 75 K and is related to OTF-2A (ref. 14). The simplest model postulates that the factors OTF-2A/OTF-2B are responsible for the B-cell-specific activity of immunoglobulin genes⁴⁻⁶. In an apparent paradox however, the octamer motif is also used by promoters and enhancers that are not lymphoid-specific¹⁵⁻¹⁸. Here, we report the isolation and initial characterization of a cDNA encoding an octamer-binding protein. This cDNA, which we will refer to

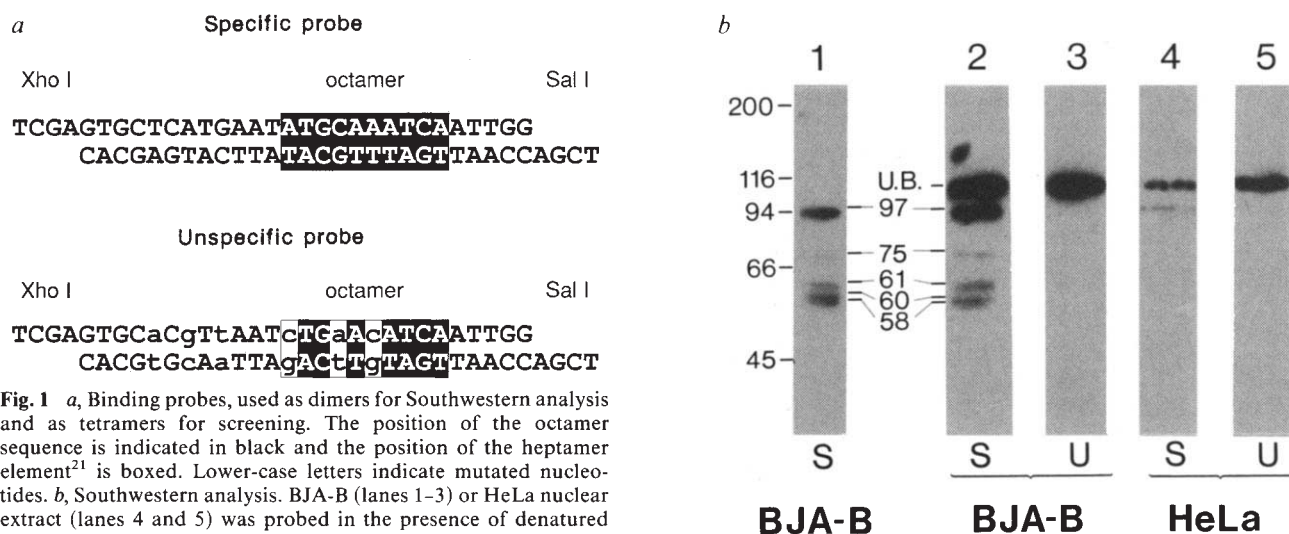


Fig. 1 *a*, Binding probes, used as dimers for Southwestern analysis and as tetramers for screening. The position of the octamer sequence is indicated in black and the position of the heptamer element²¹ is boxed. Lower-case letters indicate mutated nucleotides. *b*, Southwestern analysis. BJA-B (lanes 1–3) or HeLa nuclear extract (lanes 4 and 5) was probed in the presence of denatured calf thymus DNA (lane 1) or without carrier DNA (lanes 2–5). The sizes of protein markers (Bio-Rad) are indicated in thousands at the left. Note that the unspecific binding (U.B.) signal appears only in absence of carrier DNA. Probes: specific (S); unspecific (U).

Methods. *a*, A double-stranded oligonucleotide corresponding to the octamer-containing region of the immunoglobulin heavy-chain promoter 104.2 (ref. 45) was cloned as a *Sac*I/*Sal*I fragment with a *Xho*I site adjacent to *Sac*I into the OVEC vector⁴⁶. After one and two rounds of dimerization the resulting head to tail dimers and tetramers were released with *Xho*I and *Sal*I. The protruding 5' ends were filled in the presence of one or two [α^{32} P] NTPs for Southwestern analysis or screening, respectively. *b*, 25 μ g of nuclear extracts from BJA-B or HeLa cells¹⁴ were loaded on a 7.5% polyacrylamide/SDS gel⁴⁷ run in a PROTEAN II dual slab gel apparatus (Bio-Rad). The gel was run as described in the manual and transferred to BA85 nitrocellulose filters (Schleicher & Schuell) in running buffer containing 0.01% SDS, using a Semi-Dry Electrophoretic (JKA Biotech) at 1 mA cm⁻² for 1 h. Southwestern analysis was as described⁴⁸ with the following modifications: for pre-binding, 1% (w/v) non-fat milk powder (Carnation) was used and for binding, no milk powder was added. During binding, the specific (S) or unspecific (U) probe was present at 300,000 c.p.m. (60 fmol) ml⁻¹ with 5 μ g ml⁻¹ denatured calf thymus DNA (lane 1) or 100,000 c.p.m. (30 fmol) ml⁻¹ with no carrier DNA (lanes 2–5).

as OTF-2 cDNA, hybridizes to RNA expressed only in lymphoid cells. The predicted amino-acid sequence of this OTF-2 cDNA contains a homology to the homoeobox domain. Moreover, expression of this cDNA in HeLa cells is sufficient to activate transcription from cotransfected B-cell-specific promoters.

Library and probes

A λ GT11 cDNA expression library was prepared by oligo(dT) priming of poly(A)⁺BJA-B RNA¹⁹. This human B-cell lymphoblastoid cell line was chosen because it contains large amounts of OTF-2A protein^{6,14,20}. As a probe, we used multiple copies of a 31-base-pair (bp) DNA segment containing an octamer motif derived from the immunoglobulin heavy-chain promoter (Fig. 1*a*, top). Although it contains only one consensus octamer sequence, this 31-bp DNA segment also contains a second OTF binding site (the 'heptamer' motif^{21,22}, boxed in Fig. 1*a*; I. Kemler, E. Schreiber, M.M.M., P.M. & W.S., manuscript in preparation). The specificity of this probe was first tested by Southwestern analysis of B-cell or non-B-cell nuclear extracts. Fig. 1*b* presents the results obtained with BJA-B and HeLa extracts. In BJA-B nuclear extract, a dimer of the probe binds to OTF-1 (97 K), OTF-2A (which runs as three species with an apparent M_r of 58, 60 and 61 K¹¹), and OTF-2B (75 K), even in the presence of large amounts of carrier DNA (Fig. 1*b*, lane 1). Under less stringent conditions (without carrier DNA) an additional band at about 116 K is also detected (Fig. 1*b*, lane 2). This band represents unspecific binding (U.B.) as it is still detected with a mutated probe (Fig. 1*b*, lane 3). As expected, the only specific binding protein in nuclear extracts of HeLa cells is the ubiquitous OTF-1 (Fig. 1*b*, lanes 4 and 5).

Screening

From the unamplified λ GT11 cDNA library a total of 175,000 plaques were screened according to a modified version²³ of a previously published protocol²⁴ in which an expression library is screened with a radioactively labelled DNA binding site. In the primary screening, three recombinant phages were detected that bound specifically to the tetrameric probe, called hereafter

λ OB-A, λ OB-B and λ OB-C. Rescreening the filters with the specific and the unspecific probe (Fig. 1*a*) in parallel demonstrated that only the specific probe bound to the proteins encoded by these recombinant phages (not shown). λ OB-A, -B and -C contain inserts of about 1,750 bp, 1,900 bp and 1,925 bp, respectively. Restriction enzyme analysis revealed differences in length at the 5' and at the 3' end of the cDNAs, whereas the remainder seemed to be identical (Fig. 2*a* and Fig. 5*a*).

Binding specificity

To investigate the binding specificity of the proteins encoded by the cDNAs of λ OB-A, -B and -C, we prepared lysogens with these recombinant phages and λ GT11. Bacterial extracts were then prepared from induced lysogenic cultures and used for the bandshift experiment in Fig. 2*b*. For this assay we used a different octamer-containing probe derived from the immunoglobulin heavy-chain enhancer⁶, to verify the binding specificity with an independent probe. The binding specificity of all the bacterial extracts correlates with the pattern obtained with nuclear extracts of BJA-B cells (Fig. 2*b*, left). The complexes found with the specific probe can be competed away by an excess of unlabelled specific DNA fragment, but not by a mutated version of the same DNA fragment. Moreover, no complex appears when the mutated octamer fragment is used as probe. The bands obtained with the bacterial extracts co-migrate approximately with the complexes formed by OTF-1 and OTF-2 in BJA-B extracts, even though the bacterial protein is expressed as a β -galactosidase fusion protein. Apparently, the fusion protein was proteolytically cleaved in the bacteria, or in the extracts (our unpublished observations). Proteolysis could also account for the minor amounts of smaller complexes seen in Fig. 2.

As a further test for the binding specificity of the fusion proteins from λ OB-A and -C, we performed a methylation interference analysis with the same lysogen extracts. It had been previously shown that all the octamer-binding proteins (OTF-1, OTF-2A and 2B) produce an identical and very characteristic interference pattern independent of the cell line used for the extract^{12,14,22} (Fig. 3*b*). Figure 3*a* shows that the fusion proteins

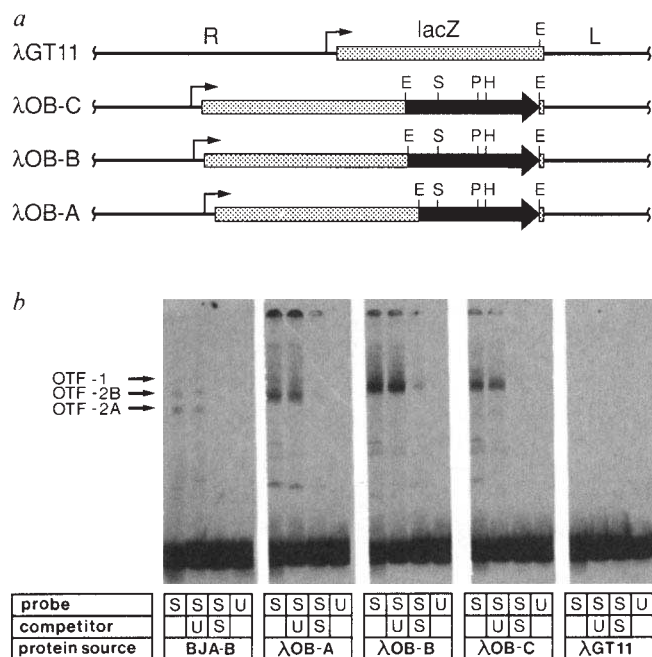


Fig. 2 Structure of recombinant phages and analysis of their fusion proteins in bandshift assays. *a*, Restriction maps with *EcoRI*(E), *SacI*(S), *PstI*(P) and *HindIII*(H). *b*, Bandshift analysis of BJA-B nuclear extract (first panel) and extracts of bacteria with the three recombinant phages of λGT11 as lysogens. The same octamer containing specific (S) and mutated unspecific (U) fragments from the immunoglobulin heavy-chain enhancer⁶ were used as probes labelled by kinase or as cold competitor fragments.

Methods. Handling of λ phages was as described⁴⁹. During the library screening²³ the only modification was the addition of 10 μg ml⁻¹ poly(dI-dC) during binding. The three positive plaques present on both duplicate filters of the first screen were confirmed and purified in one re-screen done in parallel with the specific (S) and unspecific (U) probe (see Fig. 1a). For extract preparation lysogen bacteria were prepared⁵⁰ and 100 ml cultures of each lysogen were grown at 30 °C to A₆₀₀ = 0.5, quickly brought to 43 °C and grown at 43 °C for 30 min. After addition of 0.6 ml 1M isopropylthiogalactoside the cultures were grown at 37 °C for 1 h and collected by centrifugation. Extracts were prepared as described⁵¹ (second way) and dialysed against X50 (ref. 14). For bandshifts, 2 μg BJA-B nuclear extract or 8 μg bacterial extract were incubated for 15 min on ice in 20 μl reactions containing 13,000 c.p.m. (10 fmol) probe, 8 μg poly(dIdC) 1x BEBBI (20 mM HEPES (pH 7.9), 0.5 mM EDTA, 1 mM dithiothreitol, 40 mM KCl and 4% Ficoll) and 1 pmol competitor DNA where indicated.

very closely reproduce the methylation interference pattern of the nuclear extracts (Lanes 2–5 and 7–10). The minor differences which can be observed are probably due to the bacterial part of the fusion protein.

RNA distribution

To investigate the profile of expression of OTF-2 RNA, we performed northern blot analysis of RNAs from various sources. As a probe we used the OTF-2 cDNA from λOB-C. This probe revealed a complex pattern of hybridization with RNA of lymphoid origin; in two human and in one mouse B-cell lines at least four different mRNA species exhibiting signals of different intensities were detected. The major species is approximately 6 kilobases (kb) in length and additional RNAs of 3.9, 3.3 and 2kb are also detected (Fig. 4a, lanes 5–9). These lower relative molecular mass transcripts could represent genuine mRNAs, perhaps products of alternative splicing, or specific degradation products. No hybridization could be detected with RNA extracted from four human cell lines of non-lymphoid origin (Fig. 4a,

lanes 1–4), or a human T-cell line, Hut 78 (data not shown). The absence of cross-hybridization with non-B-cell RNA suggests that the cloned cDNA encodes the lymphoid-specific OTF-2A or OTF-2B and not the ubiquitous OTF-1. Very recently others have also reported the cloning of a lymphoid specific cDNA that encodes a protein binding specifically to the octamer sequence^{20,25}. This cDNA also hybridizes to multiple transcripts, following a pattern similar to the one presented here.

Conservation

In view of its central role in immunoglobulin gene transcription we expected the gene coding for the OTF-2 protein to be conserved at least among other mammals. Genomic DNA of man, cow, mouse and rat was digested with either *EcoRI* or *HindIII* (Fig. 4b) and probed with a radioactive RNA transcript of the complete cDNA. In all cases one or two strongly hybridizing bands of comparable intensity and one to three weaker bands were detected. The weaker bands have been observed on Southern blots done with independent human placental DNA preparations. These weaker hybridizing bands resist an RNase treatment, suggesting that they indeed represent regions of very high homology to the probe. Southern blot analysis at lower stringency failed to reveal additional hybridizing fragments (data not shown). This pattern suggests that this cDNA is probably derived from a large single-copy gene or that there may be another related gene. Weak cross-hybridization (under stringent conditions) was also observed with chicken and *Drosophila* DNA (not shown).

Sequence analysis

The complete sequence of the λOB-C insert is presented in Fig. 5a. In this cDNA the predicted open reading frame starts with an ATG at position 13 and ends with a termination codon at position 1,435. The calculated *M_r* of the 474 amino-acid protein is 51 K. It therefore seems to have a lower *M_r* than the expected 60 K determined on a denaturing protein gel (see Fig. 5c, lane 2).

A homology search²⁶ through the NBRF database identified a region of strong homology between the OTF-2 cDNA and several homoeotic genes. Most interestingly, the 60 amino acids encoded from positions 886 to 1,065 of the OTF-2 cDNA show significant homology to several homoeoboxes. Figure 5b shows that OTF-2 shares up to 21 identical amino acids with the homoeoboxes of Antennapedia (*Drosophila*), Hox2.1 (mouse) and HHO.c8 (man)^{25,27,28}. The four leucine residues at positions 384, 391, 398 and 405 can be drawn as a 'leucine zipper', a newly recognized protein structural motif which is involved in protein dimerization²⁹. Furthermore, a glutamine-rich region is found in the N-terminal half of the protein (aa 75–148).

The adenosine-rich stretch at the end of the sequence, where the oligo(dT) primed cDNA synthesis started, does not seem to be a conventional poly(A) tail. The lack of a consensus polyadenylation signal and the presence of several G, C and T residues within the (A)-rich stretch indicate that the initial priming of cDNA synthesis occurred at an internal (A)-rich stretch.

To explain the discrepancy between the calculated *M_r* (51 K) and the expected *M_r* (apparent 60 K), we translated an RNA copy of the λOB-C cDNA *in vitro*. From an appropriate subclone, run-off RNA was transcribed in the correct (+) (Fig. 5c, top) and reverse (–) orientations by T7 RNA polymerase and translated in a rabbit reticulocyte lysate. The sense RNA produced a protein with an apparent *M_r* of 60 K similar to OTF-2A (Fig. 5c, lane 2), whereas the antisense RNA did not yield any detectable protein of the expected size (Fig. 5c, lane 1). Translation of the sense RNA presumably started at the first AUG which has an eight out of nine match with the ccAccATGG consensus for eukaryotic translation initiation³⁰, and ended at the stop codon at position 1,435. Although it is possible that some protein-coding information is still missing at the N terminus (and that another methionine codon is present further

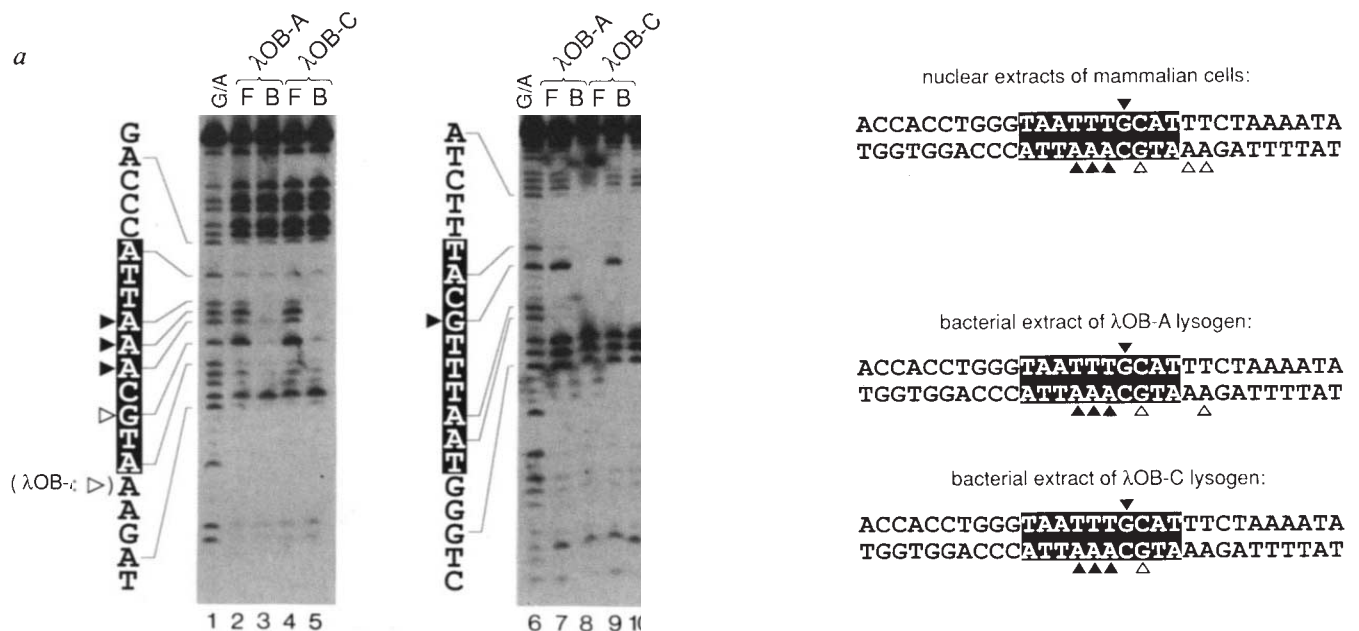


Fig. 3 Methylation interference analysis of λ OB-A and λ OB-C encoded fusion proteins. *a*, Autoradiograph of the methylation interference pattern with extracts from λ OB-A and λ OB-C lysogenic bacteria. The probe used was derived from the immunoglobulin heavy-chain enhancer and is identical to the probe used in Fig. 2, except for polylinker sequences at the ends. The nucleotides whose methylation completely or partially interferes with protein binding are indicated by solid or open triangles, respectively. Lanes labelled F correspond to the free probe (2, 4, 7, 9) and lanes labelled B correspond to the protein-bound probe (3, 5, 8, 10). *b*, Summary of the methylation interference pattern obtained on the octamer sequence with extracts from mammalian cells^{12,22} or from λ OB-A and λ OB-C lysogens.

Methods. Protein extracts were prepared as in the Fig. 2 legend, and 30 μ l (240 μ g) were used for a scaled-up binding reaction with 400,000 c.p.m. of an end-labelled, partially methylated 50-bp immunoglobulin heavy-chain enhancer fragment extending from the *Dde*I site to the *Hin*I site (reclaimed by *Xho*I or *Sal*I cleavage of a pUC subclone, P. M. unpublished). After preparative band-shift, the bound and free DNA were isolated from the polyacrylamide gel and purified by passing over a DEAE column. After elution the DNA was alkali-cleaved as in a G > A reaction⁵² and the products were separated on an 8% polyacrylamide/8M urea sequencing gel.

upstream), we believe that very little—if anything—is lacking, as in gel retardation assays the *in vitro*-translated OTF-2 protein exactly co-migrates with OTF-2A from BJA-B nuclear extract (not shown).

Transcriptional activation

The protein encoded by the cDNA of λ OB-C has the apparent M_r and binding specificity expected of OTF-2A. As the octamer motif is sufficient to confer B-cell-specific transcription³⁻⁶, we anticipated that expression of the cloned OTF-2 cDNA might be sufficient to activate a B-cell-specific promoter in non-B cells. To test this hypothesis we inserted the cDNA of λ OB-C in both orientations into a eukaryotic expression vector, such that OTF-2 expression was under the control of the strong cytomegalovirus (CMV) enhancer/promoter³¹ (Fig. 6a). As target genes, we used a series of plasmids that contain synthetic promoters comprising the β -globin TATA box and various other binding sites for transcription factors. All these plasmids carry an SV40 enhancer downstream of the globin gene (Fig. 6b, lines 1-5). As the SV40 enhancer is active in a variety of cell lines^{32,33}, any differences in transcriptional efficiency should reflect the activity of the respective promoter in a particular cell line.

All five constructs were analysed by transfecting them individually into BJA-B cells (to show the B-cell-specific activity of the octamer constructs), or together with the OTF-2 expression vector into HeLa cells (to test transactivation by OTF-2). When the promoter contained only a TATA box, transcripts were hardly detectable in BJA-B or HeLa cells due to the very weak activity of this incomplete promoter (Fig. 6c, lanes 1 and 6). Addition of an Sp1 binding site (Fig. 6b, line 2) led to strong signals in both cell lines (Fig. 6c, lanes 2 and 9), as expected from the ubiquitous presence of the Sp1 protein³⁴. The signal remained unchanged upon co-transfection with either of

the two OTF-2 expression vectors (Fig. 6c, lanes 9-11). This control shows that, under the experimental conditions used, OTF-2 does not have a general effect on transcription. A different picture emerged when the same experiment was done with two B-cell-specific promoters containing an octamer motif (Fig. 6b, lines 3 and 5). As expected both promoters were efficiently transcribed in BJA-B cells (Fig. 6c, lanes 3 and 5), whereas very few transcripts were detected in HeLa cells (Fig. 6c, lanes 12 and 18). Most interestingly, both promoters could be activated by cotransfection of the correctly expressed OTF-2 cDNA, but not by cotransfection of the antisense cDNA (Fig. 6c, lanes 12-14 and 18-20).

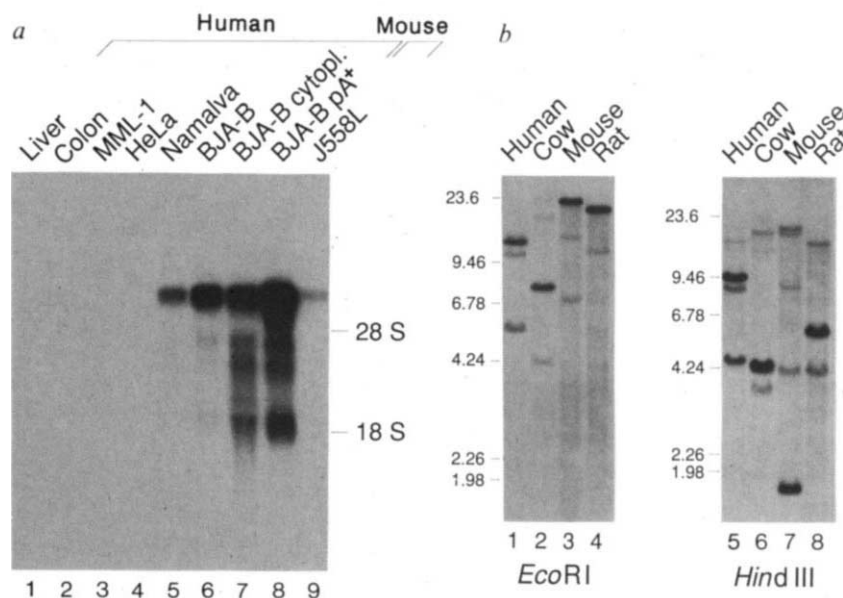
Subsequently the octamer motif in construct OCTA(2) was altered by three point mutations (Fig. 6b, line 4) known to abolish binding of OTF proteins (ref. 6 and Fig. 2b). This mutated promoter did not function in BJA-B cells (Fig. 6c, lane 4) nor was it stimulated by cotransfection with the OTF-2 cDNA expression vector (Fig. 6c, lanes 15-17). As only the octamer binding site was mutated, it is clearly this element which is responsible for OTF-2-dependent activation.

Discussion

Characteristics of the cloned OTF-2 cDNA. An OTF-2 cDNA was isolated by screening a B-cell cDNA expression library with a radioactively labelled binding site derived from an immunoglobulin heavy-chain promoter. Northern blot analysis revealed that OTF-2 mRNA is restricted to lymphoid cells, probably to the B lineage, as no RNA was detected in one human T-cell line (Hut 78, data not shown). In the positive cell lines several RNAs hybridize specifically with the OTF-2 cDNA. The major RNA, approximately 6 kb long, and also the 3.9 and 3.3 kb species are significantly larger than would be expected to be necessary for translation of OTF-2A. These RNAs may

Fig. 4 Messenger RNA and gene structure of OTF-2. *a*, Northern blot analysis of OTF-2 RNA from various organs or cell lines, as indicated. The origin of the cell lines is as follows: MML1 is derived from a human melanoma, HeLa from a cervix carcinoma, Namalva from a Burkitt lymphoma, BJA-B is a lymphoblastoid cell line and J558L is a mouse myeloma. The positions of 18S and 28S ribosomal RNAs are indicated. *b*, Southern blot analysis of genomic DNA from various mammals. As size markers the positions of *Hind*III fragments of λ DNA are indicated in kb.

Methods. For Northern analysis, total RNA (5 μ g, lanes 1–6; 9), cytoplasmic RNA (5 μ g, lane 7) or poly(A)⁺ RNA (0.5 μ g, lane 8) was denatured and separated on a 1.1% agarose-formaldehyde gel and blotted onto Genescreen (Dupont). After blotting, the filter was baked for 30 min at 80 °C and UV cross-linked for 30 s. UV shadowing of the filter showed that equivalent amounts of RNA had been loaded in the slots, not shown. Prehybridization was at 65 °C in 50% formamide 5 $\times$ SSC, 50 mM sodium phosphate pH 6.5, 8 $\times$ Denhardt's, 0.5 mg ml⁻¹ yeast RNA, 1% SDS. Hybridization was carried out in the same solution containing 15 μ g ml⁻¹ poly(C) and 10⁷ c.p.m. ml⁻¹ of an antisense RNA probe complementary to the entire λ OB-C insert. The filters were washed as follows: 2 $\times$ SSC, 1% SDS, 65 °C, 30 min; 0.2 $\times$ SSC, 1% SDS, 68 °C, 4 $\times$ 30 min; 2 $\times$ SSC, 20 °C, 3 $\times$ 5 min; 2 $\times$ SSC, 1 mg ml⁻¹ RNase A, 20 °C, 20 min; 2 $\times$ SSC, 0.5% SDS, 65 °C, 15 min. For Southern blot analysis, 10 μ g DNA from human placenta, mouse liver, calf thymus or the rat hepatoma FTO-2b cell line was digested to completion and separated on a 0.8% agarose gel run in Tris-acetate buffer. The gel was alkali blotted onto Genescreen and cross-linked as indicated above. Hybridization⁵³ was in 0.5 M sodium phosphate pH 7.2, 7% SDS, 1 mM EDTA, 1% bovine serum albumin at 68 °C with 10⁷ c.p.m. ml⁻¹ of an antisense RNA probe as above. Filters were washed in 40 mM sodium phosphate, 1% SDS at 65–69 °C, 4 $\times$ 30 min, followed by an RNaseA wash as for the northern analysis.



represent products of differential splicing and may have a coding capacity for other related proteins, such as OTF-2B. At present, we do not know from which of these RNAs the isolated cDNA was derived. Primer extension experiments and the isolation of further cDNAs will be necessary to determine the precise relationship between these RNA species.

By Southern blot analysis of genomic DNA, the OTF-2 gene seems to be a large single-copy gene which is very highly conserved, at least amongst mammals.

The largest OTF-2 cDNA isolated, λ OB-C, is 1.95 kb long and contains a 1.45-kb open reading frame. When transcribed and translated *in vitro*, this cDNA produces a protein with an apparent M_r of about 60 K, the same as that of OTF-2A. This *in vitro*-translated protein and the fusion protein present in extracts from induced λ OB-C lysogens both have a DNA binding specificity that is virtually identical to that of octamer-binding proteins from BJA-B cells.

Structure of OTF-2, a homoeotic protein. The most striking feature revealed by the OTF-2 cDNA sequence is a significant homology, between amino acids 292 and 351, to the homoeobox, a highly conserved stretch of 60 amino acids shared by many genes controlling development (see ref. 35 for review). For example, there are 21 identical amino acids between OTF-2 and the Antp, Dfd or Hox2.1 homoeoboxes (Fig. 5b). This is only slightly less than the 24 out of 60 identical amino acids shared between Antp and H2.0, the product of a newly identified *Drosophila* gene which contains the most divergent homoeobox so far characterised in metazoa³⁶. Interestingly, H2.0 is the first *Drosophila* homoeobox gene isolated which exhibits a tissue-specific pattern of expression³⁶.

The homoeobox encodes a helix-turn-helix motif³⁷ (W.J. Gehring and K. Wüthrich, personal communication, see Fig. 5b) and has been shown to be a DNA-binding domain with loose specificity^{38,39}. It is thus reasonable to assume that the OTF-2 homoeobox is at least part of the DNA-binding domain of that protein. Preliminary band-shift experiments with truncated *in vitro*-translated OTF-2 proteins suggest that this is indeed the case (data not shown and ref. 27). It seems unlikely

that the OTF-2 gene is a homoeotic gene in the true sense. Instead it looks as though the helix-turn-helix structure may be a DNA-binding motif more generally used among higher eukaryotes than has been recognized so far. Very recently, others have noticed that the diverged homoeobox present in OTF-2 is highly conserved among three other genes, and that the region of shared homology extends further towards the N terminus of these proteins, thereby defining a new conserved motif^{25,40,41}.

OTF-2 may also contain a 'leucine zipper'²⁹ formed by the leucine residues at positions 384, 391, 398 and 405 (with the fourth helical turn interrupted by a proline at position 407). Whether this leucine zipper is used for potential dimerization of the OTF-2 protein remains to be established.

Moreover, the N-terminal half of the protein contains a glutamine-rich region which may be involved in transcriptional activation, as it seems to be the case for Sp1 (R. Tjian, personal communication).

Specific transcriptional activation by cloned OTF-2. We have shown that expression of the OTF-2 cDNA in HeLa cells was sufficient to strongly activate cotransfected B-cell-specific promoters. This *trans*-activation was specific and dependent on the presence of an intact octamer motif in the promoter of the reporter gene. No *trans*-activation was observed with promoters containing only a TATA box, a TATA box and an Sp1 binding site or a TATA box and a mutated octamer motif.

As not only the immunoglobulin promoters, but also the immunoglobulin heavy-chain enhancer contain an octamer motif, we performed OTF-2 expression vector co-transfections with target plasmids in which we replaced the downstream SV40 enhancer by the immunoglobulin heavy-chain enhancer or by multimers of the octamer motif (essentially as in ref. 6). Although these constructs were very active in BJA-B cells, they could not be stimulated in HeLa cells by OTF-2 expression (data not shown). Several explanations of this result are possible. One is that the cloned OTF-2 can only act on the promoter, and that a different OTF-2-related factor is responsible for activation from a downstream position. As already suggested elsewhere¹⁴, the second lymphoid-specific octamer factor, OTF-2B, is a

c

octa-C ⊕

ATG

cDNA

5' 3'

116 94 66 45

octa-C

− +

~60

1 2

b

1 10 20 30 40 50 60

Helix 1 Helix 2 Helix 3

Antp RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIETAHAHCLTSDRQ-IKIWFQNRNMKWKKEN

Dfd PKRCRTATYTRHQILELEKEFHFNRYLTRRRRIETAHTVLSDRQ-IKIWFQNRNMKWKKDN

BSH4 QRRSRTFTAEQLEALBRAFSRTQYEDVVTREELAQTALTAR-IQVWFSNRRARLRKHS

H2.0 RSWSRAVFSNLRKGLIEIQQQKYITKPDRRKLAARNLTDQAQ-VKVFQNRNMKWRHTR

Hox 2.1 GKRAATATYTRYQTLELEKEFHFNRYLTRRRRIETAHAHCLSDRQ-IKIWFQNRNMKWKKDN

Hox 1.1 RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIETAHAHCLTSDRQ-IKIWFQNRNMKWKKEH

HHC.c8 RRRGRQIYSRYQTLELEKEFHFNRYLTRRRRIETANADCLTSDRQ-IKIWFQNRNMKWKKES

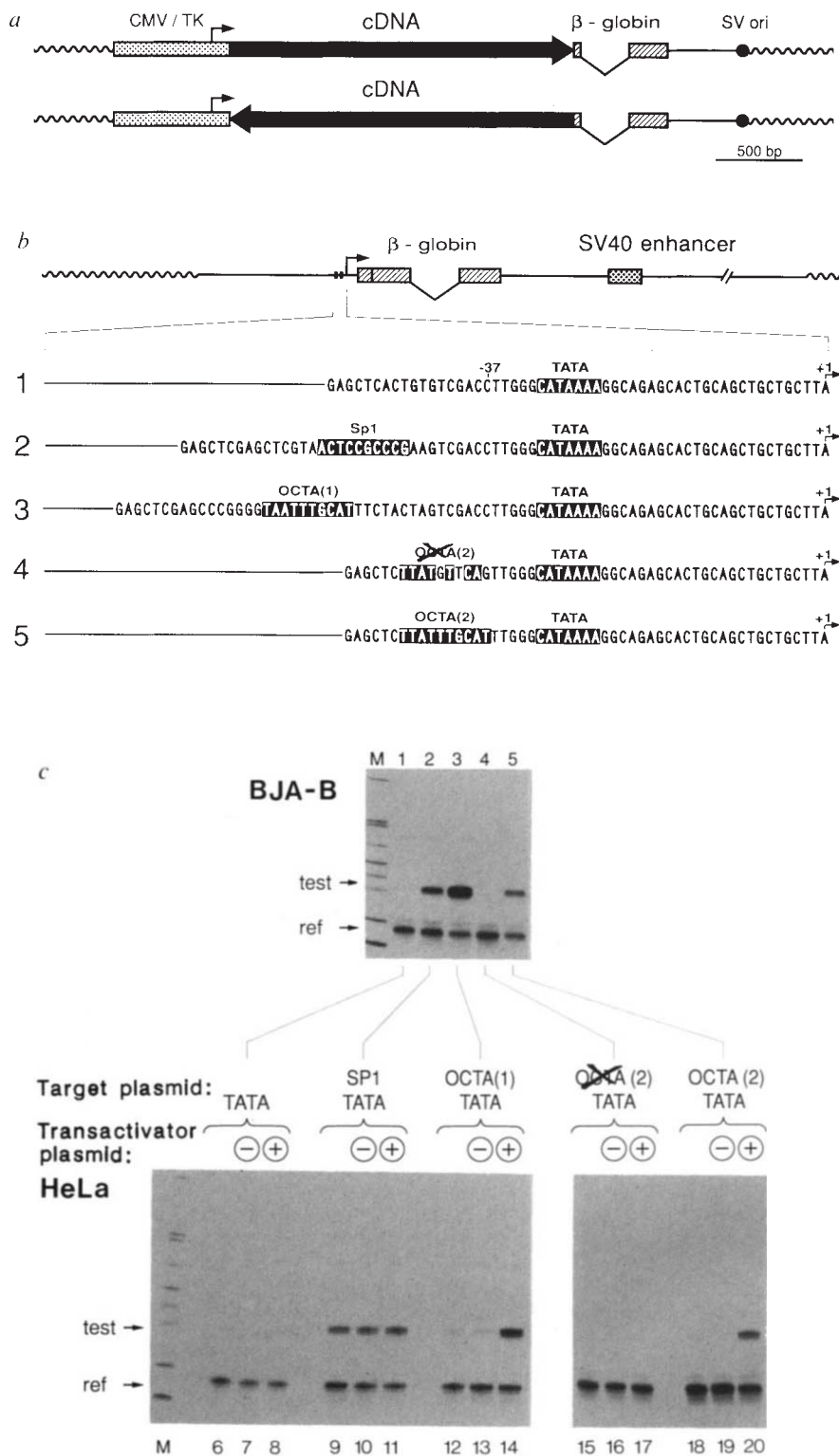
Homeo Consensus RKRGRATYTRYQTLELEKEFHFNRYLTRRRRIETAHAHCLTSDRQ-IKIWFQNRNMKWKKEN

OTF-2 RRRKRTSIETNVRFALEKSLANOKETSEEILLTAEQNHM-SKEVIRVWFQNRNRQKEKRIN

Methods. The cDNA of λ OB-C was inserted in both orientations into the Bluescript SK vector (Stratagene). After linearization with *Kpn*I, nested deletions were made by exonuclease III, mung bean nuclease and T4 DNA polymerase treatment before religation and *Escherichia coli* transformation⁵⁴. Clones with deletions of the appropriate size were identified by restriction analysis and used for dideoxy sequencing (Kilobase kit, BRL) after alkali denaturation of the DNA. Most of the sequence was determined in both orientations twice with both normal and modified (deazaGTP or dITP) nucleotides. Gaps or ambiguities were resolved by using internal primers and by chemical sequencing⁵² of some fragments. The ends of the λ OB-A and λ OB-B cDNAs were determined from pUC subclones. The sequence data were compiled using the DNAsar or the StadenPlus programs (Amersham). For the translation experiment the *Eco*R1 insert of λ OB-C was cloned in both orientations into the *Eco*R1 site of pGemini1 (Promega) to yield pGemC+ and pGemC-, respectively. 1 μ g of both *Xba*I-linearized templates was transcribed in a 30 μ l reaction containing 1 mM of all rNTPs, 10 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, 6 mM dithiothreitol, 40 units RNasin (Promega) and 20 units T7 RNA polymerase (New England Biolabs) at 37 °C for 40 min. Translation reactions were as described⁵⁵ with 4 μ l RNA/DNA mixture in a final volume of 20 μ l; 2 μ l of each reaction were electrophoresed as described in Fig. 1.

Fig. 6 Specific promoter activation by cloned OTF-2 cDNA. *a*, The expression vectors pOEVI (+) or (-) express the OTF-2 cDNA in sense (+) or antisense (-) orientation from the cytomegalovirus enhancer/promoter and herpes simplex virus thymidine kinase leader as in vector pSTC-X556 (ref. 31). *b*, Reporter plasmids were based on the OVEC vector⁴⁶ and contained various factor binding sites upstream of the β -globin TATA box; all constructs contained an SV40 enhancer 3' to the β -globin gene. Construct 1 is OVEC, construct 2 (Sp1) contains an Sp1 site derived from the metallothionein gene promoter⁵⁶, construct 3 (OCTA 1) construct an octamer motif derived from the immunoglobulin heavy-chain enhancer (T. Gerster, unpublished results). Construct 4 contains an octamer motif derived from the histone H2B promoter with three point mutations which abolish factor binding and activity, and construct 5 (OCTA 2) contains the same octamer motif intact (P.M., unpublished results). The positions of the various factor binding sites is indicated by a black box. *c*, Activity of the various reporter constructs was monitored by RNase mapping of RNA from transfected cells. The upper panel (lanes 1-5) shows a transfection experiment in B cells and demonstrates that the octamer containing reporter plasmids are active in B cells. The lower panel (lanes 6-20) shows a *trans*-activation experiment in HeLa cells. Together with the indicated reporter plasmid, calf thymus DNA (lanes 6, 9, 12, 15, 18), antisense *trans*-activator (lanes 7, 10, 13, 16, 19) or sense *trans*-activator plasmid DNA (lanes 8, 11, 14, 17, 20) was cotransfected. 'Ref' indicates the position of the reference signal (produced by the plasmid OVEC-Ref⁴⁶) and 'test' indicates the position of the correctly initiated RNA derived from the reporter plasmids.

Methods. All cloning was done by standard procedures³⁷ by inserting gel-purified double-stranded oligonucleotides into the appropriate restriction sites of the OVEC vector. All clones were verified by DNA sequencing. BJA-B cells were transfected by the DEAE-dextran procedure as described⁶. HeLa cells were transfected by the CaPO₄ method with 15 μ g reporter plasmid, 2 μ g reference plasmid (OVEC-Ref) and 3 μ g of either *trans*-activator plasmid or calf thymus DNA. 12-14 h after DNA addition the CaPO₄-DNA precipitate was removed and cells were shocked with dimethylsulphoxide (final concentration 25%). Cells were then incubated for about 30 h and cytoplasmic RNA was extracted. 20 μ g RNA was used for hybridization to a radioactive complementary-strand RNA probe (spanning positions -37 to +179 of the vector). Hybridization products were digested with RNase A (6.5 μ g ml⁻¹ Sigma) and RNase T1 (10 units ml⁻¹, Boehringer) at 37°C for 60 min and separated on a 6% polyacrylamide/8M urea gel.



candidate for this putative enhancer activation factor.

At present, we cannot exclude the possibility that the activation of B-cell-specific promoters in HeLa cells could also be achieved by overexpression of the ubiquitous OTF-1 factor. We consider this possibility unlikely however, because of the previously established lymphoid specificity of these promoters^{4,5}, which correlates with the lymphoid specific expression of OTF-2^{11,12,20,42}.

How specific is the OTF-2/DNA interaction? The heptamer sequence CTCATGA (see Fig. 1a) is a recently recognized conserved motif found in immunoglobulin heavy-chain pro-

moters, which has been shown to be important for immunoglobulin promoter function²¹. Although the heptamer exhibits no obvious homology to the octamer motif, it also is a binding site for OTF-1 (see ref. 22) and for the OTF-2 protein encoded by the cDNA described here (I. Kemler, E. Shreiber, M.M.M., P.M. & W.S., in preparation). As shown here, OTF-2 specifically *trans*-activates octamer-containing target plasmids. Furthermore, this cloned OTF-2 cDNA is also able to specifically *trans*-activate at least two different target sequences for homoeotic proteins⁴³. These, and other observations⁴⁴, indicate that the octamer factor(s) can interact specifically with a number

of different DNA sequences which deviate from the TNATTTGCAT consensus. A major challenge will be to analyse how the binding specificity is brought about.

Finally, it will be particularly interesting to see by which control mechanism the OTF-2 gene is regulated. Is the regulation at the transcriptional level (perhaps via autoregulation by the OTF-2 gene product itself?) or at the post-transcriptional level? In either case, the cloned OTF-2 cDNA will be instrumental in elucidating the mechanisms which control immunoglobulin gene expression.

Received 7 October; accepted 31 October 1988.

1. Falkner, F. G., Mocikat, R. & Zachau, H. G. *Nucleic Acids Research* **14**, 8819-8827 (1986).
2. Falkner, F. G. & Zachau, H. G. *Nature* **310**, 71-74 (1984).
3. Parslow, T. G., Blair, D. L., Murphy, W. J. & Granner, D. K. *Proc. natn. Acad. Sci. USA* **81**, 2650-2654 (1984).
4. Wirth, T., Staudt, L. & Baltimore, D. *Nature* **329**, 174-178 (1987).
5. Dreyfus, M., Doyen, N. & Rougeon, F. *EMBO J* **6**, 1685-1690 (1987).
6. Gerster, T., Matthias, P., Thali, M., Jiricny, J. & Schaffner, W. *EMBO J* **6**, 1323-1330 (1987).
7. Singh, H., Sen, R., Baltimore, D. & Sharp, P. A. *Nature* **319**, 154-158 (1986).
8. Sturm, R., Baumruker, T., Franza Jr, B. R. & Herr, W. *Genes Dev.* **1**, 1147-1160 (1987).
9. O'Neill, E. A. *et al. Science* **241**, 1210-1213 (1988).
10. Fletcher, C., Heintz, N. & Roeder, R. G. *Cell* **51**, 773-781 (1987).
11. Scheidereit, C., Heguy, A. & Roeder, R. G. *Cell* **51**, 783-793 (1987).
12. Staudt, L. M. *et al. Nature* **323**, 640-643 (1986).
13. Landolfi, N. F., Capra, J. D. & Tucker, P. W. *Nature* **323**, 548-551 (1986).
14. Schreiber, E., Matthias, P., Müller, M. M. & Schaffner, W. *EMBO J* (in the press).
15. Mattaj, J. W., Lienhard, S., Jiricny, J. & De Robertis, E. M. *Nature* **316**, 163-167 (1985).
16. Zenke, M. *et al. EMBO J* **5**, 387-397 (1986).
17. Barberis, A., Superti-Furga, G. & Busslinger, M. *Cell* **50**, 347-359 (1987).
18. Rosales, R. *et al. EMBO J* **6**, 3015-3025 (1987).
19. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
20. Staudt, L. M. *et al. Science* **241**, 577-580 (1988).
21. Eaton, S. & Calame, K. *Proc. natn. Acad. Sci. USA* **84**, 7634-7638 (1987).
22. Pruijn, G. J. M., van Driel, W., van Miltenburg, R. T. & van der Vliet, P. C. *EMBO J* **6**, 3771-3778 (1987).
23. Vinson, C. R., LaMarco, K. L., Johnson, P. F., Landschulz, W. H. & McKnight, S. L. *Genes Dev.* **2**, 801-806 (1988).
24. Singh, J. H., LeBowitz, A. S., Baldwin, A. S. & Sharp, P. A. *Cell* **52**, 415-423 (1988).
25. Clerc, R. C., Corcoran, L. M., LeBowitz, J. H., Baltimore, D. & Sharp, P. A. *Genes Dev.* (in the press).
26. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. USA* **80**, 726-730 (1983).
27. Ko, H., Fast, P., McBridge, W. & Staudt, L. M. *Cell* **55**, 135-144 (1988).
28. Bürglin, T. R. *Cell* **53**, 339-340 (1988).
29. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240**, 1759-1764 (1988).

We thank Rudi Mészleányi for excellent technical assistance, Sandro Rusconi for experimental advice and help with *in vitro* translations; Hugh R. B. Pelham, Walter Gehring, Edgar Schreiber, François Meyer and Roberto Cattaneo for comments or gifts of material; Hugh R. B. Pelham, Christopher Southgate and Deborah Maguire for critical reading of the manuscript, Günther Schütz for hospitality to one of us (P.M.) during the initial phase of this work and Fritz Ochsenbein for artwork. This work was supported by the Swiss NSF and the Kanton of Zürich.

30. Kozak, M. *Nucleic Acids Res.* **15**, 8125-8148 (1987).
31. Severne, Y., Wieland, S., Schaffner, W. & Rusconi, S. *EMBO J* **7**, 2503-2508 (1988).
32. Schirm, S., Jiricny, J. & Schaffner, W. *Genes Dev.* **1**, 65-74 (1987).
33. Ondek, B., Shepard, A. & Herr, W. *EMBO J* **6**, 1017-1025 (1987).
34. Kadonaga, J. T. & Tjian, R. *Proc. natn. Acad. Sci. USA* **83**, 5889-5893 (1986).
35. Gehring, W. J. *Science* **236**, 1245-1252 (1987).
36. Barad, M., Jack, T., Chadwick, R. & McGinnis, W. *EMBO J* **7**, 2151-2161 (1988).
37. Laughon, A. & Scott, M. P. *Nature* **310**, 25-31 (1984).
38. Desplan, C., Theis, J. & O'Farrell, P. H. *Nature* **318**, 630-635 (1985).
39. Hoey, T. & Levine, M. *Nature* **332**, 858-861 (1988).
40. Herr, W. *et al. Genes Dev.* (in the press).
41. Sturm, R. A., Das, G. & Herr, W. *Genes Dev.* (in the press).
42. LeBowitz, J. H., Kobayashi, T., Staudt, L., Baltimore, D. & Sharp, P. A. *Genes Dev.* **2**, 1227-1237 (1988).
43. Thali, M., Müller, M. M., DeLorenzi, M., Matthias, P. & Bienz, M. *Nature* **336**, 598-601 (1988).
44. Baumruker, T., Sturm, R. & Herr, W. *Genes Dev.* (in the press).
45. Ballard, D. W. & Bothwell, A. *Proc. natn. Acad. Sci. USA* **83**, 9626-9630 (1986).
46. Westin, G., Gerster, T., Müller, M. M., Schaffner, G. & Schaffner, W. *Nucleic Acids Res.* **15**, 6787-6798 (1987).
47. Laemmli, U. K. *Nature* **227**, 680-685 (1970).
48. Miskimins, W. K., Roberts, M. P., McClelland, A. & Ruddle, F. H. *Proc. natn. Acad. Sci. USA* **82**, 6741-6744 (1985).
49. Huynh, T. V., Young, R. A. & Davis, R. W. *DNA Cloning Techniques: A Practical Approach* (ed. Glover, D.) 49-78 (IRL, Oxford, 1985).
50. Hynes, R. O., Schwarzbauer, J. E. & Tamkun, J. W. *Methods in Enzym.* **144**, 447-463 (1987).
51. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. *Genes Dev.* **2**, 786-800 (1988).
52. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. USA* **74**, 560-564 (1977).
53. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. USA* **81**, 1991-1995 (1984).
54. Henikoff, S. *Gene* **28**, 351-359 (1984).
55. Rusconi, S. & Yamamoto, K. R. *EMBO J* **6**, 1309-1315 (1987).
56. Höller, M., Westin, G., Jiricny, J. & Schaffner, W. *Genes Dev.* **2**, 1127-1135 (1988).
57. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning—A Laboratory Manual* (Cold Spring Harbor, New York, 1982).

A human lymphoid-specific transcription factor that activates immunoglobulin genes is a homoeobox protein

Claus Scheidereit*, James A. Cromlish, Thomas Gerster, Kiyoshi Kawakami*, Carmen-Gloria Balmaceda, R. Alexander Currie & Robert G. Roeder

Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, New York, New York 10021, USA

The human lymphoid-specific transcription factor OTF-2 contains a homoeodomain that is required for DNA binding and binds specifically to DNA elements that are recognized by Drosophila homoeodomain proteins, suggesting coevolutionary relationships between mammalian and invertebrate homoeodomain proteins and their DNA recognition elements.

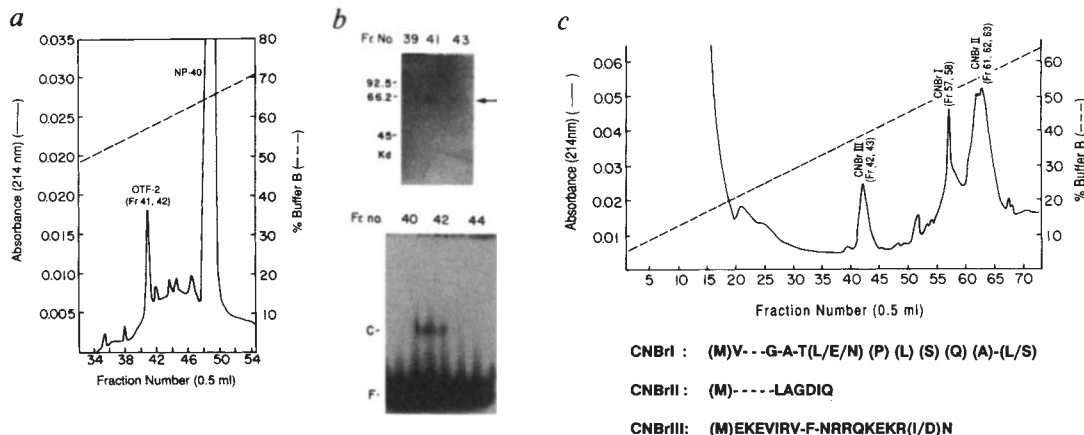
THE conserved octamer sequence motif (ATGCAAAT) (refs 1-3) has been found in a wide variety of promoters and enhancers. In the immunoglobulin genes it appears to be the chief determinant for the B-cell specificity of both light and heavy chain promoters⁴⁻⁶ and the heavy chain enhancer^{7,8}. These elements are recognized by the B cell-specific octamer-binding transcription factor OTF-2, which was the first tissue-specific gene regulatory factor identified^{9,10} and purified to apparent

homogeneity in a transcriptionally active form¹¹. The octamer element is also an important functional element of many ubiquitously expressed genes and, for instance, is known to mediate the S phase activation of the histone H2B gene¹². These elements are recognized by one or more ubiquitous octamer binding proteins (NFIII (ref. 13), NF-A1 (ref. 14), OTF-1 (ref. 15) and OBP100 (ref. 16)). Despite the presumed functional differences between the lymphoid-specific and the ubiquitous protein(s) the site-specific binding properties of these proteins are very similar or identical^{10,11}, which might reflect homologies in the DNA-binding domains of the proteins.

Structural studies on both general and inducible vertebrate

* Present addresses: Otto Warburg Laboratory of Molecular Biology, Max-Planck-Institute for Molecular Genetics, Ihnestrasse 73, 1000 Berlin 33, FRG (C.S.) and Department of Biology, Jichi Medical School, Yakushiji, Minamikawachi-machi, Kawachi-gun, Tochigi-ken 329-04, Japan (K.K.).

Fig. 1 HPLC chromatography, renaturation and protein-sequencing of OTF-2. **a**, Gradient elution profile (25 to 100% solvent B over 30 min) of OTF-2 by reverse phase HPLC. The N-terminal amino-acid analysis of OTF-2 yielded the following sequence: V H (P/S) S M G A P K I (R/L) M--P L, where positions obtained at low sequencing yields are indicated within parentheses and unknown amino acids with dashes. **b** (upper panel), SDS-gel electrophoresis with Coomassie brilliant blue staining of HPLC purified OTF-2. The positions of protein molecular weight markers and OTF-2 are indicated. **b** (lower panel), Gel retardation analysis of renatured HPLC OTF-2 fractions. C and F are OTF-2 complexed and free DNA respectively. **c** (upper panel), Gradient elution profile (5 to 80% solvent B over 30 min) of OTF-2 cyanogen bromide cleavage products fractionated by reverse phase HPLC. The elution position of peptide peaks used for amino-acid sequence analysis are indicated as CNBrI-III with their derived amino acids displayed below (**c**, lower panel). A minor peptide fragment in CNBrIII yielded the following sequence: (M)--P L R A E K Q G L D S P S.



Methods. Reverse phase HPLC was performed on a Vydac 214TP54 protein C4 column (0.46 × 25 cm) in which the elution system consisted of a linear gradient of acetonitrile in the presence of trifluoroacetic acid. Solvent A was 0.1% (v/v) trifluoroacetic acid in water and solvent B was 0.1% (v/v) trifluoroacetic acid in acetonitrile/water (9:1) (v/v). All elutions were performed at a flow rate of 1.5 ml min⁻¹ at room temperature. For the renaturation experiment, ~20 pmoles of affinity-purified OTF-2 (ref. 11) was fractionated by reverse phase HPLC and 0.2 ml aliquots were dried and subjected to SDS-polyacrylamide gel electrophoresis. The remainder of the fractions (0.3 ml) was dried, dissolved in guanidine hydrochloride and diluted¹¹. Gel retardation analysis was performed¹¹ on 2% (15 μl) of this material. N-terminal analysis was performed on 100 pmoles of HPLC purified OTF-2. For the generation of OTF-2 CNBr fragments, about 200 pmoles of OTF-2 was subjected to reverse phase HPLC, and the resulting pool concentrated to a volume of 20 μl. Cyanogen bromide digestion used 5 mg ml⁻¹ CNBr in degassed 70% (v/v) formic acid in the dark for 24 h at room temperature in a final volume of 100 μl. Subsequently the reaction mixture was diluted to 500 μl with buffer A, and the reaction products were separated by reverse phase HPLC. Fractions containing the CNBr fragments gave initial sequencing yields of ~50 pmoles when sequenced on an Applied Biosystems Model 470 A gas phase sequencer equipped with a Model 120 A phenylthiohydantoin analyser.

transcription factors, including steroid receptors (reviewed in ref. 17), TFIIIA (ref. 18), SP1 (ref. 19), AP1/c-jun (ref. 20) and C/EBP (ref. 21), have indicated the presence of conserved domains. These include the zinc finger and the leucine zipper motifs (reviewed in ref. 22) that appear to be directly or indirectly involved in DNA-binding. The homoeodomain is another DNA-binding structure which was originally recognized in regulatory proteins from invertebrates (reviewed in ref. 23). It contains a helix-turn-helix motif similar to that found in the DNA-binding domains of procaryotic regulatory proteins²⁴.

As a first step towards understanding the structural, evolutionary and regulatory relationships of octamer-binding proteins we have cloned cDNAs encoding the entire OTF-2 and show that this factor is also a homoeodomain-containing protein.

Isolation of OTF-2 cDNA clones

Transcriptionally active OTF-2 was purified to apparent homogeneity by affinity chromatography¹¹ and then applied to a reversed phase HPLC column (Fig. 1a). It was eluted reproducibly in a single major peak (fractions 41 and 42) which contained both the previously described¹¹ subspecies of relative molecular mass 58–62K and the octamer-specific DNA binding activity of OTF-2 (upper and lower panels, respectively, of Fig. 1b).

After cyanogen bromide cleavage of HPLC-purified OTF-2, the major peptides (CNBr-I, -II and -III) were resolved by reversed-phase HPLC and sequenced (Fig. 1c). The sequence of CNBr-III, which contained the most conserved part of the homoeodomain (see below), was used to synthesize both degenerate and 'best-guess' oligonucleotides. Screening of a cDNA library from Namalwa cells, a human Burkitt lymphoma line, yielded several clones that hybridized to these probes and that on sequence analysis proved to share a common internal sequence encoding CNBr-III (details in legend to Fig. 2).

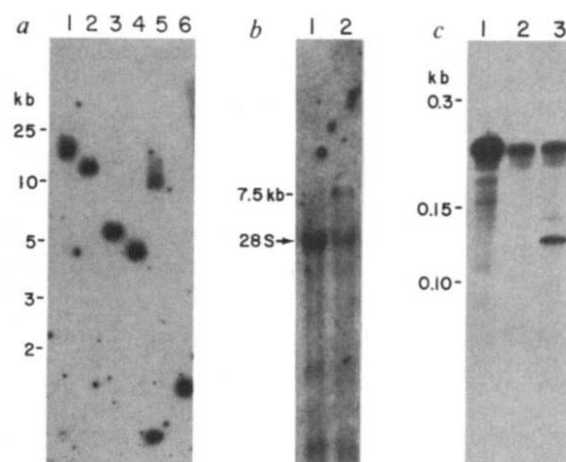
OTF-2 contains a homoeodomain

Figure 2 shows the combined cDNA sequence of overlapping OTF-2 clones 32 and 15, which together form a 1,389-base pair open reading frame encoding a 463-amino acid protein of 49.5K. The occurrence of the N-terminal sequence (Fig. 1a legend), an overlapping CNBr sequence (Fig. 1c legend), and CNBrI-III sequences within the predicted amino-acid sequence (underlined residues) proves that the clones represent OTF-2. The position of the N-terminal peptide sequence argues strongly that translation begins with the methionine at nucleotide position 1. The apparently greater size of natural OTF-2 on SDS gels presumably reflects an abnormal electrophoretic mobility.

The N-terminal region between amino acids 74 and 189 is particularly rich in glutamine, proline and leucine (20% each), whereas another serine-rich region (20%) is found between amino acids 221 and 271. Clusters of glutamines or serines and threonines have also been described for the transcription factor SP1 (ref. 19). The region between amino acids 281 to 340 contains a homoeodomain which includes the CNBr-III sequence. The homoeodomain, considered further below, is itself part of a longer region that is relatively rich in charged amino acids. Near the carboxy terminus (between amino acids 356 and 450) is another stretch rich in serine and threonine (19% and 13%, respectively). This region contains (between amino acids 373 and 394) a leucine zipper-like structure²² with four leucines (at seven-residue intervals) in the described periodic array. Compared with the previously described zipper domains²², it is unusual in its amino-acid content, containing only one charged amino acid. It could nonetheless form an amphipathic α helix, with eleven serine and threonine residues localized on the side opposite the leucine array. Although the functional significance of this zipper, say for dimerization²², remains to be proven, OTF-2 has been shown to bind as a dimer to two distinct functional elements (heptamer and octamer) in the

Fig. 3 Southern and Northern blot analysis and RNase mapping. *a*, Southern blot: Namalwa DNA was digested with *Asp*718, *Bam*H1, *Eco*R1, *Hind*III, *Pst*I and *Pvu*II (lanes 1–6). DNA molecular weight marker positions are indicated. *b*, Northern blot: Hela (lane 1) and Namalwa mRNA (lane 2) was analysed using a *Pst*I fragment from clone 15 (Fig. 2 legend) as probe. The size of the hybridizing message is indicated. The nonspecific signal in each lane probably represents cross-hybridizing residual 28S RNA. *c*, RNase A mapping: probe alone (lane 1) and protected transcripts in cytoplasmic Hela RNA (lane 2) and in Namalwa RNA (lane 3). Molecular weight marker positions are indicated. The faint signal above the main band might reflect a minor RNA species containing the divergent 5' sequences of clone 15.

Methods. For the Southern analysis *a*, genomic DNA was digested with restriction enzymes, electrophoresed on 1% agarose gels and transferred to NYTRAN nylon membranes (Schleicher and Schuell). Filters were washed with $0.05 \times$ SSC, 0.5% SDS at 65 °C. The probe was a random primer-labelled *Hind*III fragment from clone 15 extending from position 1,124 to the right end in Fig. 2. For Northern blotting *b*, total RNA was isolated from Hela and Namalwa cells and poly(A)⁺ RNA purified over oligo(dT) cellulose. 5 µg samples were separated on 1% agarose gels containing 0.66M formaldehyde and transferred to NYTRAN membranes. A random primer labelled fragment containing sequences 865 to 1,481 (Fig. 2) was used as probe. Final washes were carried out with $0.1 \times$ or $0.2 \times$ SSPE ($1 \times$ SSPE is 0.18 M NaCl, 10 mM NaH₂PO₄, pH 7.4, 1 mM EDTA), 0.1% SDS at 60 °C. For the RNase mapping *c*, a *Hind*III fragment from clone 15 containing sequences 167 to 282 (Fig. 2) as well as divergent 5' sequences, was transcribed by T7 polymerase on the lower strand, using [α -³²P]GTP. This probe was hybridized to 50 µg cytoplasmic Hela or Namalwa RNA and digested with RNase A. RNase-resistant molecules were analysed on a 6.5% denaturing polyacrylamide gel.



(lanes 1–9) were compared in a gel shift assay with an octamer probe. Without competitors the expressed protein generated a complex (lane 1, lower arrow) whose faster migration relative to the complex generated by OTF-2 (lane 11, upper arrow) was consistent with the presence of a protein of ~40K. For each protein-DNA complex, there was strong competition with histone H2B and heavy-chain promoter octamer elements (lanes 2, 5, 12 and 15) and significant, but slightly reduced, competition with heavy-chain enhancer, kappa light-chain promoter and herpes simplex virus ICP0 octamer elements (lanes 3, 4, 8, 13, 14 and 18). In contrast, no competition was observed with non-related histone H4 promoter elements or a mutant ICP0 octamer element (lanes 6, 7, 9, 16, 17 and 19).

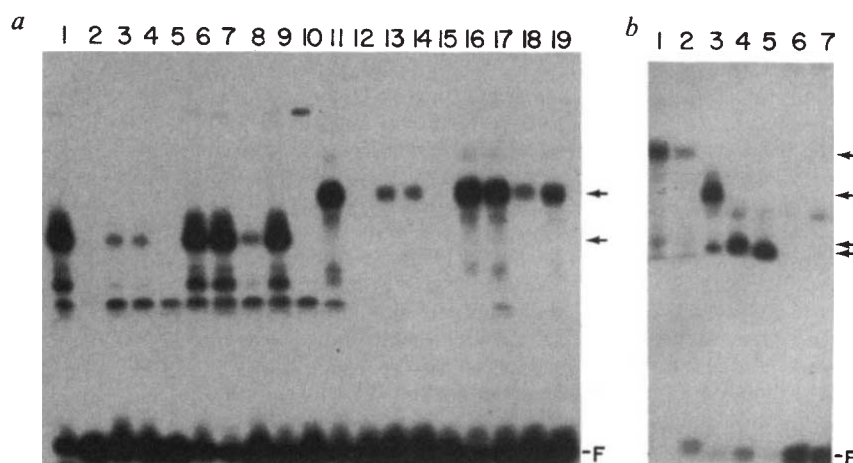
Importantly, both OTF-2 and the cDNA-encoded protein showed the same difference in affinities for the various competitors—further evidence for identity between the DNA-binding domains of these proteins. Similarly, like OTF-2, the clone 15-derived protein has been shown to form the heptamer-octamer dependent dimeric complex on a heavy-chain promoter fragment (L. Poellinger and R.G.R., in preparation).

To more precisely localize the DNA-binding domain, differentially truncated transcripts from clone 15 were translated and analysed by a gel shift assay (Fig. 4b). Products containing the protein sequence extending only to amino acid 374 (lane 4, third arrow from top) or 359 (lane 5, fourth arrow) bound to the H2B promoter octamer with an affinity similar to that shown by the non-truncated clone 15-derived protein (lane 3, second arrow) or by purified OTF-2 (lanes 1 and 2, first arrow). In contrast, truncation at amino acid 211 led to a complete loss of binding (lane 6), indicating that the 148 amino acid homoeodomain-containing region between the *Hind*III and *Nco*I sites is required for DNA binding. Along with the following studies, these results indicate that the homoeodomain is involved in site-specific binding by OTF-2. They also indicate that the potential leucine zipper is not essential for binding of OTF-2 to an isolated octamer element.

Binding to 'engrailed' consensus sequence

Homoeoproteins that bind to specific DNA sequences include *engrailed* (*en*), *fushi tarazu* (*ftz*), *even-skipped* (*eve*), *ultra-*

Fig. 4 *In vitro* expression of OTF-2 cDNA and subregions. *a*, Expression of clone 15 in rabbit reticulocyte lysate and comparison with OTF-2. Gel retardation assay with an H2B probe: lanes 1–9: *in vitro* translated OTF-2; lane 10: reticulocyte lysate programmed with a T7 RNA polymerase-generated antistrand RNA from clone 15 (negative control); lanes 11–19: purified OTF-2. Arrows point to the specific protein-DNA complexes and F denotes free DNA. No competitor added (lanes 1, 11); octamer-containing or control oligonucleotides added: histone H2B (ref. 15) (lanes 2, 12); heavy chain enhancer (IgH position 526–555) (lanes 3, 13); kappa light-chain promoter¹¹ (lanes 4, 14); heavy chain promoter (lanes 5, 15); histone H4 promoter (lanes 6, 16, 7, 17); herpes simplex virus TAATGARAT (ref. 31) (lanes 8, 9); TAATGARAT triple mutant (ref. 31) (lanes 9, 19). *b*, Expression of subregions of clone 15 in reticulocyte lysates. Gel retardation assay using purified OTF-2 (lanes 1 and 2) and *in vitro* translated proteins of RNAs derived from cDNA templates truncated at *Eco*RV, *Hind*III, *Nde*I and *Nco*I sites (lanes 3–6), or *in vitro* translated brome mosaic virus RNA (lane 7). Arrows denote protein-DNA complexes, F is free DNA.



Methods. Clone 15 DNA (1 µg) was linearized at various 3' restriction sites and transcribed by T3 polymerase (Stratagene). The RNA (1 µg) was translated in a rabbit reticulocyte lysate system (Promega). 1–2 µl was used for gel shift analysis¹¹. Affinity-purified OTF-2 (~1 ng per assay) was used for the gel retardation analysis in parallel. Competitor oligonucleotides were used at 170-fold molar excess over the probe.

bithorax (*ubx*) and yeast mating type proteins²⁶⁻²⁸. The recognition elements show a remarkable degree of similarity: *en* protein binds to *en* promoter elements (TCAATTAAAT consensus), which are also recognized by *ftz* and *eve* proteins, whereas *eve* protein also recognizes different *eve* promoter elements (TCAG-CACCG consensus). A (TAA)_n-type sequence in the *ubx* and *antp* promoters is in turn bound by *ubx* and, albeit weakly, by *en*^{29,30} proteins. The yeast mating type α -2 protein binds to palindromic sequences²⁸ which closely resemble the octamer element ATGCAAAT, with 5, 6 or 7 out of 8 matches in each axis. Except for *eve* recognition sites, the other sites mentioned also show a certain resemblance to the octamer element. It is noteworthy that different types of OTF-recognition sites have been found; these include the classical octamer (above), the NFIII site¹³, the SV40 octamer¹⁶ and the TAATGARAT (ref. 31) element. Therefore we analysed the possibility of a shared DNA-binding specificity between OTF-2 and the *Drosophila* proteins experimentally.

The binding of purified OTF-2 to a histone H2B octamer-containing DNA fragment (Fig. 5a, lane 1), was blocked by both an immunoglobulin octamer-containing oligonucleotide (lanes 6 and 7) and (somewhat less effectively) an oligonucleotide containing three copies of the *engrailed* consensus sequence (lanes 2 and 3); binding was not affected by the same *engrailed* oligonucleotide with consensus region mutations (lanes 4 and 5) which eliminate binding of the *engrailed* protein (Y. Ohkuma and C. Desplan, personal communication), or by a non-related oligonucleotide containing a binding site for the transcription factor NF κ B (lanes 8 and 9).

In addition, OTF-2 formed a complex with an *engrailed* DNA probe (Fig. 5b); specificity was indicated by competition with kappa octamer sites (lanes 9-12) and, somewhat less effectively, with *engrailed* sites (lanes 1-4), but not with the mutant *engrailed* (lanes 5-8) or non-related NF κ B (lanes 13-16) oligonucleotides. The probable OTF-2 binding sites within the *engrailed* probe are one or more motifs related to the octamer sequences ATGCAAAT or ATGATAAT; each of the latter is bound by OTF-2 equally well (data not shown). With 5 to 6 out of 8 matches, these motifs start at positions 10, 16, 22, 28 and 34 in the *engrailed* oligonucleotide sequence (Fig. 5 legend).

Conversely, other studies have shown that expressed *engrailed* protein when bound specifically to the *engrailed* element is competed out by octamer and *engrailed* oligonucleotides, but not by the NF κ B oligonucleotide (Y. Ohkuma and C. Desplan, personal communication) and that a highly purified *antennapedia* homoeodomain binds specifically to a vertebrate octamer element (W. Gehring, personal communication). Thus, conservation of the homoeodomain (see below) between various homoeoproteins appears to determine a very similar DNA binding specificity. Taken together these observations suggest a strong evolutionary conservation not only of the homoeodomain, but also of the respective DNA targets.

Homology with other homoeobox proteins

A comparison of the OTF-2 homoeodomain with a selection of other homoeobox proteins is shown in Fig. 6. There is a high degree of homology (identical amino acids and conservative changes) both in the carboxy-terminal region that contains the proposed helix 2 and helix 3 regions (see below) and in additional N-terminal positions, most notably in positions that are generally well-conserved between other homoeobox proteins. The OTF-2 homoeodomain is related to the *antennapedia* domain with 33% identity.

Based on certain amino-acid conservations with the helix-turn-helix motif of the prokaryotic regulators (reviewed in ref. 32), this structure has been proposed as the main DNA-interaction structure in homoeodomains^{33,34}. In agreement with this suggestion, OTF-2 also reveals characteristic hydrophobic amino acids at positions 34, 38 and 48, as well as alanine at position 35 and isoleucine at position 45.

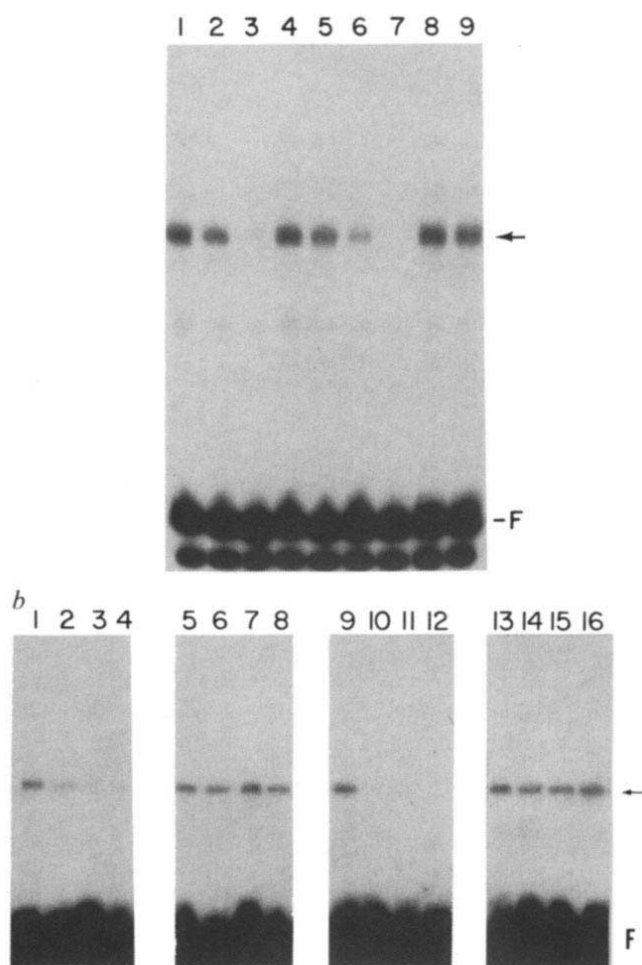


Fig. 5 Specific binding of OTF-2 to an '*engrailed*' consensus sequence. **a**, Competition for H2B octamer-OTF-2 complexes. No competitors (lane 1) or 100-fold (lanes 2, 4, 6, 8) and 1,000-fold (lanes 3, 5, 7, 9) molar excess of *engrailed* binding sites (lanes 2, 3), mutated *engrailed* sites (lanes 4, 5), kappa promoter octamer (lanes 6, 7) or NF κ B oligonucleotides⁴³ (lanes 8, 9). **b**, Competition for *engrailed* site-OTF-2 complexes. Zero, 10-, 30- and 100-fold molar excess of *engrailed* sites (lanes 1-4), mutated *engrailed* sites (lanes 5-8), immunoglobulin kappa promoter (lanes 9-12) or NF κ B oligonucleotides (lanes 13-16) over *engrailed* probe was used.

Methods. Gel retardation was performed essentially as described¹¹ using a 70-base pair H2B probe or the *engrailed* probe²⁶ containing the region CGGATCCTCAATTAAATGATCAA-TTAAATGATCAATTAAATGA. Mutations (transversions) introduced into the mutant *engrailed* sites are underlined.

Conclusions and perspectives

The data presented document the isolation of a cDNA encoding OTF-2, a lymphoid-specific transcription factor that appears to have a key role in determining the cell type specificity of immunoglobulin genes. We show that OTF-2 is encoded by a single-copy gene and, importantly, that its cell type specificity is regulated at the level of mRNA abundance. Further studies of the developmental regulation of the OTF-2 gene are now possible and offer the promise of novel information regarding the signals and pathways involved in the commitment, differentiation and function of B cells. A more immediate question is whether OTF-2 is itself the sole determinant for the B cell-specific function of immunoglobulin octamer elements, or whether this involves still other lymphoid-specific factors that interact either with OTF-2 or with the general transcription machinery (such as the TATA factor).

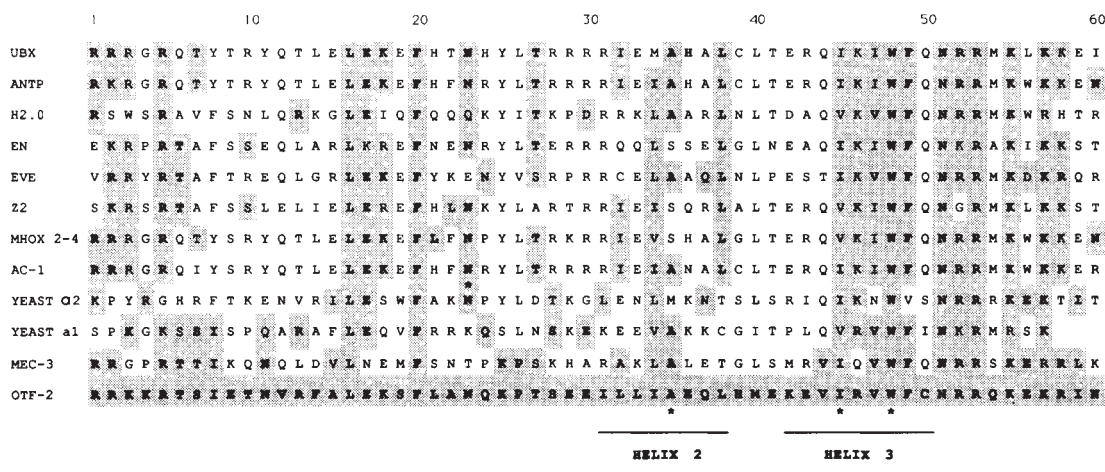


Fig. 6 Alignment of the homoeodomain sequence of OTF-2 with several other homoeodomains. Conserved amino acids are shaded following the scheme (K,R), (E,D), (Q,N), (P,G), (L,V,I,A), (Y,W,F,H), (M,C) and (S,T), and identical amino acids are printed in bold letters. The helix-turn-helix motif has also been discussed as a structural feature of the homoeodomain³³ and is diagrammed below the sequence. The asterisk in the α -2 sequence indicates a 3-amino acid deletion for maximum match, asterisks below the OTF sequence mark conserved positions in the helix-turn-helix motif. Literature sources in refs 37 and 39.

In addition, the availability of a cloned cDNA permits more detailed investigations of structural determinants important both for sequence-specific DNA recognition and for transcriptional activation, including the mechanisms by which general transcription factors are stimulated (compare ref. 35) and the basis for the functional distinctions between OTF-2 and other factors (such as OTF-1) with similar or identical binding specificities. An important basis of these analyses is the availability of cell-free systems^{4,11} in which OTF-2 function can readily be studied.

A surprising but significant result has been the finding that OTF-2 is a homoeobox protein, making it the first mammalian homoeobox protein for which a specific function has been demonstrated. This function as a transcriptional activator is in agreement with recent studies of yeast and *Drosophila* homoeobox proteins (ref. 28; J. Jaines and P. H. O'Farrell, manuscript in preparation) and with expectations regarding other mammalian homoeobox containing proteins³⁶. Our studies also suggest, in accordance with the studies in *Drosophila* and yeast, that the mammalian homoeodomain is involved in site-specific DNA recognition.

A number of homoeobox proteins are involved in segmentation processes in early insect development (reviewed in ref. 23). However, their occurrence in specialized tissues in both invertebrates and vertebrates, as well as direct genetic analysis, suggest a broader role in the differentiation and function of specialized cells (reviewed in ref. 37). The results presented here also indicate the direct involvement of mammalian homoeobox proteins in specialized cell function, although an earlier role in the differentiation process cannot be excluded. Studies in *Drosophila* have indicated that a given site can be recognized by distinct homoeobox proteins, and that a given homoeobox protein can recognize more than one site^{26,27}, and have led to the proposal of regulatory interactions between homoeobox genes, with the activity of a given gene being determined by the relative affinities and concentrations of the interacting factors. An alternative model, involving DNA sequence-dependent combinatorial interactions of distinct homoeobox proteins, has been suggested for the yeast α -1 and α -2 mating-type proteins³⁸.

It is not yet clear whether similar mechanisms will apply to OTF-2 (and other mammalian homoeobox proteins) or whether OTF-2 will have a single specificity, with target gene activity regulated simply by the presence of OTF-2. Related to these considerations, OTF-2 and OTF-1 (below) have very similar or identical DNA recognition specificities and interact cooperatively, as homodimers or heterodimers, through distinct sequence elements in the heavy chain promoter (L. Poellinger

and R.G.R., manuscript in preparation). Thus, although it seems likely that the homoeobox proteins are structurally and evolutionarily related, it is not yet clear, at least for transcription factors like OTF-2, whether they have evolved independently and with clearly distinct functions, or whether they have evolved as an interacting multigene family.

A related question is whether the functions of homoeobox proteins in metazoans will necessarily be restricted to developmentally regulated genes. The mammalian OTF-1 has a binding specificity (on purified DNA) indistinguishable from that of OTF-2, but is a ubiquitous factor that is involved in transcription of a variety of ubiquitously expressed genes. If OTF-1 proves to have a DNA-binding domain related to that of OTF-2, as seems probable, then a more generalized role for homoeodomains will be indicated and we may expect to see their occurrence in other general factors.

We thank Drs Alessandra Pierani and Lorenz Poellinger for help and discussion, Steve Kay and Kevin Gorman for data bank searches and computer graphics, respectively, and Claude Desplan for the gift of *engrailed* oligonucleotides. We thank Donna Atherton at the Rockefeller University Protein Sequence Facility for performing amino-acid sequence analysis. Furthermore we thank Michael M. Müller and Drs Patrick Matthias and Walter Schaffner for suggesting to us a possible frameshift at position 1226, experimentally verified by us, in the original version of our sequence. The protein sequence facility is supported in part by a grant from the National Institutes of Health. This work was supported by a grant from the National Cancer Institute to R.G.R. and by general support from the Pew Trusts to The Rockefeller University. J.C. was supported by a Centennial Fellowship from the Medical Research Council of Canada; T.G. was supported by a grant from the Schweizerischer Nationalfonds and C.S. by grants from the Deutsche Forschungsgemeinschaft and Deutscher Akademischer Austauschdienst DAAD, Sonderprogramm Gentechnologie.

Received 7 October; accepted 28 October 1988.

1. Harvey, R. P., Robins, A. J. & Wells, J. R. E. *Nucleic Acids Res.* **10**, 7851-7863 (1982).
2. Falkner, F. G. & Zachau, H. G. *Nature* **310**, 71-74 (1984).
3. Parslow, T. G., Blair, D. L., Murphy, W. J. & Granner, D. K. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2650-2654 (1984).
4. Mizushima-Sugano, J. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8511-8515 (1986).
5. Wirth, T., Staudt, L. & Baltimore, D. *Nature* **329**, 174-178 (1987).
6. Dreyfus, M., Doyen, N. & Rougeon, F. *EMBO J.* **6**, 1685-1690 (1987).
7. Gerster, T., Matthias, P., Thali, M., Jiricny, J. & Schaffner, W. *EMBO J.* **6**, 1323-1330 (1987).
8. Lenardo, M., Pierce, J. W. & Baltimore, D. *Science* **236**, 1573-1577 (1987).
9. Landolfi, N. F., Capra, J. D. & Tucker, P. W. *Nature* **323**, 548-551 (1986).
10. Staudt, L. M. *et al.* *Nature* **323**, 640-643 (1986).
11. Scheidereit, C., Heguy, A. & Roeder, R. G. *Cell* **51**, 783-793 (1987).

12. LaBella, F., Sive, H. L., Roeder, R. G. & Heintz, N. *Genes Dev.* **2**, 32–39 (1988).
13. Pruijn, G. J. M., van Driel, W. & van der Vliet, P. C. *Nature* **322**, 656–659 (1986).
14. Singh, H., Sen, R., Baltimore, D. & Sharp, P. *Nature* **319**, 154–157 (1986).
15. Fletcher, C., Heintz, N. & Roeder, R. G. *Cell* **51**, 773–781 (1987).
16. Sturm, R., Baumruker, B. R., Franza, J. R. & Herr, W. *Genes Dev.* **1**, 1147–1160 (1987).
17. Evans, R. M. *Science* **240**, 889–895 (1988).
18. Ginsberg, A. M., King, B. O. & Roeder, R. G. *Cell* **39**, 479–489 (1984).
19. Kadonaga, J. T., Carner, K. R., Masiaz, F. R. & Tjian, R. *Cell* **51**, 1079–1090 (1987).
20. Bohmann, D. *et al.* *Science* **238**, 1386–1392 (1987).
21. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. & McKnight, S. L. *Genes Dev.* **2**, 786–800 (1988a).
22. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. *Science* **240**, 1759–1764 (1988b).
23. Gehring, W. *Science* **236**, 1245–1252 (1987).
24. Schleif, R. *Science* **241**, 1182–1187 (1988).
25. Staudt, L. M. *et al.* *Science* **241**, 577–580 (1988).
26. Desplan, C., Theis, J. & O'Farrell, P. H. *Cell* **54**, 1081–1090 (1988).
27. Hoey, T. & Levine, M. *Nature* **332**, 858–861 (1988).
28. Johnson, A. D. & Herskowitz, I. *Cell* **42**, 237–247 (1985).
29. Beachy, P., Krasnow, M., Gavis, E. & Hogness, D. *Cell* (in the press).
30. Robertson, M. *Nature* **327**, 556–557 (1987).
31. Gerster, T. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 6347–51 (1988).
32. Pabo, C. O. & Sauer, R. T. *A. Rev. Biochem.* **53**, 293–321 (1984).
33. Laughon, A. & Scott, M. *Nature* **310**, 25–31 (1984).
34. Shepherd, J. W. C., McGinnis, W., Carrasco, A. E., De Robertis, E. M. & Gehring, W. J. *Nature* **310**, 70–71 (1984).
35. Horikoshi, M., Hai, T., Lin, Y. S., Green, M. R. & Roeder, R. G. *Cell* **54**, 1033–1042 (1988).
36. Dressler, G. R. & Gruss, P. *Trends Genet.* **4**, 214–219 (1988).
37. Way, J. C. & Chalfie, M. *Cell* **54**, 5–16 (1988).
38. Miller, A. M., MacKay, V. L. & Nasmyth, K. *Nature* **314**, 598–603 (1985).
39. Barad, M., Jack, T., Chadwick, R. & McGinnis, W. *EMBO J.* **7**, 2151–2161 (1988).
40. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
41. Lathé, R. *J. molec. Biol.* **183**, 1–12 (1985).
42. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
43. Kawakami, K., Scheidereit, C. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4700–4704 (1988).

LETTERS TO NATURE

Stellar archaeology and black widow pulsars

David Eichler

Department of Physics, Ben-Gurion University, PO Box 653, 84105 Beer Sheva, Israel, and Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742, USA

Many forms of dark matter might collect in or near stars. Although it may have little observable effect on the star itself, one way to find such dark matter would be to eliminate the rest of the star. The recent discovery¹ of PSR1957+20 suggests the existence of 'black widow' neutron stars, which destroy their companion by ablation, and supports previous conjectures² that isolated millisecond pulsars were spun up by companions that were later disposed of. Any exotic, non-degradable residuum would remain in orbit around the pulsar. In this letter we note that a statistically viable sample of millisecond pulsars may present a means of looking for excavated exotic matter at the cores of previously existing stars. The mass threshold of detectability using pulsar timing, $3 \times 10^{-8} M_{\odot}$ or less (J. Taylor, personal communication), would provide a much stronger constraint than previous considerations³ and is of astrophysical interest.

It has been suggested previously (ref. 4 and references therein) that shadow matter—that is, an unfamiliar sector of matter that interacts with itself by means of 'subatomic' forces and with familiar matter only by gravitation—might exist and collect in planets or stars during their formation. Charged quark nuggets, and possibly neutral ones, could couple to ordinary matter sufficiently to be included in normal compact objects. It has been suggested that mini-black holes exist in stellar cores for various reasons. They could originate primordially or from stellar capture of bosonic WIMPs (weakly interacting massive particles).

It has also been proposed that the 160-min solar oscillation is driven by a small black hole⁵ or a shadow planet⁶ in a 2×10^4 -km orbit around the Sun's centre. The analysis in ref. 6 allows a mass of $\sim 10^{-7} M_{\odot}$, which causes a precession of the perihelion of Mercury of $0.02''$ per century, a radius of $r \approx 10^8$ cm and an orbital dissipation rate of $\sim 10^{24}$ erg s⁻¹, allowing the orbit to survive for $\sim 10^{17}$ s. For our purposes, the exotic matter is just as interesting if it has settled to the centre of the Sun.

Another interesting possibility for interactive dark matter is bosonic stars, a stable configuration of a self-repelling bosonic field ϕ (for example, self-repelling by a ϕ^4 term in the lagrangian) contained by gravitation (M. Colpi, S. L. Shapiro and I. Wasserman, personal communication). The masses of such objects depend on the parameters in the field theory. For sufficiently weak self-repulsion, they would merely fuse into black holes if collected by stellar cores.

The growth time τ of black holes, if they accrete at the Eddington limit, is of the order of the Salpeter time,

$(\epsilon \sigma_T c / 4\pi G m_p)$, where ϵ is the efficiency for converting rest mass into energy and is probably of the order of 0.1 for black holes. Here σ_T is the Thomson scattering cross-section, m_p is the proton mass, and c and G have their usual meanings. By the Dirac large-number coincidence, τ is of the order of $\epsilon H_0^{-1} / 4\pi$ (H_0 is the Hubble constant). This is considerably smaller than the Hubble time H_0^{-1} , and the black hole would disrupt the familiar evolution of a main-sequence star when its own luminosity exceeds that expected from nuclear burning. For completeness, we consider the finely tuned case where the growth time is of the order of $H_0^{-1} / 30$. (This is less than an order of magnitude slower than the above estimate. No *a priori* justification is given, other than the fact that it is close to that obtained above.) Black holes starting from sufficiently small masses might not typically have affected stellar evolution yet, but could have grown (by e^x , where $x \approx 20$, say) to masses that might be detectable as companions to pulsars. If we take this mass to be $\sim 10^{-7} M_{\odot}$, the accretion luminosity is $\sim 10^{30}$ – 10^{31} erg s⁻¹. It is not clear that this can be ruled out by observations of stellar populations, even if it occurs in a significant fraction of old stars. Moreover, capture of the seed objects might be biased to the more massive stars. By virtue of the larger luminosities and shorter lifetimes of massive stars, small black holes within them would be more innocuous than in low-mass stars.

It may also be that dark matter capture, which typically has a strong velocity dependence, occurs preferentially in close binaries (orbital velocities ~ 200 km s⁻¹) where the relative velocity between the stars and the dark matter can be significantly perturbed in either direction by the binary motion and three-body gravitational effects. Dark matter that could be captured only by close binaries would end up in the ideal location for the detection scheme proposed here.

To summarize, some forms of dark matter could collect in stellar cores and settle into a sufficiently compact form that it would survive disruption of the host stars.

The primary evidence that the companion to PSR1957+20 is being ablated is that the emission at 430 MHz is eclipsed sharply at a radius of $\sim 5 \times 10^{10}$ cm. Assuming the plasma frequency there exceeds 430 MHz, and assuming that the outflow velocity exceeds the escape velocity, $\sim 10^7$ cm s⁻¹, one infers a mass loss rate of at least $\sim 10^{15}$ g s⁻¹. Given the estimated mass of 5×10^{31} g, the evaporation time of the companion is at most ~ 1 to 2×10^9 yr. The current lower limit (J. Taylor, personal communication) to the spin-down time of the pulsar is 5×10^8 yr, so the observations can reasonably be interpreted in terms of a pulsar in the process of vaporizing its companion to completion.

The means by which PSR1957+20 ablates its companion is not understood. In any ablation scenario, however, if the companion is ablated to below the mass of minimum radius (0.05 – $0.1 M_{\odot}$), it is likely to go to completion.

Because exotic matter, even if it exists, does not necessarily exist in all stellar systems, a statistically viable sample of solitary millisecond pulsars would be necessary to determine that exotic

residua are not to be found in nature. Indeed, the first several millisecond pulsars found are either solitary, to within experimental limits, or have binary companions that are consistent with conventional evolutionary scenarios.

But the recently reported discovery⁸ of PSR0021-72, if real, defies a conventional explanation unless it is being observed nearly face on. Having a period of 32 min, an eccentricity of 0.33, and projected semimajor axis $a \sin(i) = 585$ km, the pulsar must have a companion that is both compact, $R \ll 10^9$ cm, and light, $M \ll M_\odot$, unless $\sin(i) \ll 1$. The compactness, even for a very light companion, follows from the failure of tidal forces to circularize the orbit, and for $\sin(i) \sim 1$, this argument alone rules out a very light dwarf which is composed of normal material.

The most conventional explanation for this binary, from the point of view of fundamental physics, is that it consists of two neutron stars in orbit around each other, with the orbital plane viewed nearly face on, $\sin(i) < 3 \times 10^{-3}$. The chance probability for this to occur, leaving aside possible selection effects in the search procedure, is less than 10^{-5} , but it should not be dismissed in view of the radical nature of the alternatives. If several such sources are discovered, however, such that it becomes impossible to attribute the observation to improbable inclination angles, then a radical interpretation will seemingly be inescapable.

Although the evidence is insufficient to prove that PSR0021-72 has an exotic companion, this case nevertheless illustrates how observations of millisecond pulsars would be capable of revealing dark matter, if indeed it exists. If pulsars evaporate their companions, leaving behind exotic matter remnants in places

where they can be detected, then a survey of pulsars should be capable of finding the remnants, or of determining that they do not exist.

There should be at least $\sim 1,000$ millisecond pulsars in the Galaxy, given that most of those already discovered lie in a very small patch of the sky (several square degrees), and more than that if the known pulsars are mostly close and hence distributed isotropically. With such a large sample set, it should be possible to overcome statistically the uncertainties that apply to individual cases, and thus to find whether a variety of low-mass companions in various stages of disappearance is needed to explain the observed distribution of orbital parameters. If evidence of this sort is found, a search for exotic residua that remain as ultra-light companions to solitary millisecond pulsars would yield information on the dynamical properties of dark matter, thereby constraining a variety of dark-matter cosmologies.

I thank Drs J. Bekenstein, P. Joss and D. Stinebring for valuable suggestions. This research was supported in part by the NSF.

Received 12 September; accepted 27 October 1988.

1. Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. *Nature* **333**, 237–239 (1988).
2. Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. *Nature* **300**, 728–730 (1982).
3. Kolb, E. W., Seckel, D. & Turner, M. S. *Nature* **314**, 415–419 (1984).
4. Okum, L. B. *Zh. eksp. teor. Fiz.* **79**, 694–699 (1980).
5. Severny, A. B., Kotov, V. A. & Tsap, T. T. *Astr. Zh.* **56**, 137–141 (1979).
6. Blinnikov, S. I. & Khlopov, M. Yu. *Sol. Phys.* **82**, 383–385 (1983).
7. Eichler, D. & Levinson, A. *Astrophys. J. Lett.* (in the press).
8. Ables, J. G. *et al.* *IAU Circ.* no. 4602 (1988).

Low-energy Galactic-Centre γ -rays from low-mass X-ray binaries

W. Kluźniak, M. Ruderman, J. Shaham & M. Tavani

Department of Physics and Astrophysics Laboratory,
Columbia University, New York 10027, USA

The hard X-ray and low-energy γ -ray emission from the Galactic Centre region (GCR) has four components^{1,2}: a power-law continuum between 20/50 keV and 200/300 keV with inverse power-law photon index α in the range ~ 2.5 to ~ 3.1 ; a harder spectrum between 200/300 keV and ~ 511 keV with $\alpha \approx 1.0$ – 1.5 ; a narrow electron–positron annihilation line at 511 keV, reported to disappear^{3,4} in $< 1/2$ yr, although the temporal variation is controversial⁵; and an equally variable continuum emission between 511 keV and several MeV ('MeV bump'). All four have luminosities of 10^{37} – 10^{38} erg s⁻¹, if the source is located 10 kpc away. We propose non-thermal processes in low-mass X-ray binaries concentrated in the galactic bulge as the direct source of the three continuum components of the emission and also, possibly, as the indirect source of the 511-keV line. We suggest that the softer power-law component of the GCR continuum arises from synchrotron emission of relativistic electrons in the strongly non-uniform magnetic field of the neutron star, and, more tentatively, that the MeV bump is the result of interaction of harder γ -rays with the power-law photons. The hardest power law may be due to Compton scattering of relativistic electrons or photons.

Massive-black-hole models^{6–8} of GCR hard radiation may be in conflict with recent coded-aperture observations^{9,10}. The total X-ray (0.5–30 keV) luminosity of the Sgr A complex is $< 10^{36}$ erg s⁻¹. The only bright hard X-ray (> 30 keV) source close by is ~ 50 arcmin away from the Galactic-Centre direction and even this source, 1E1740.7–2942, with a luminosity of 2.4×10^{37} erg s⁻¹ (at 10 kpc), contributes only little to the HEAO-3 (35° field of view) spectral flux. An $\sim 10^2 M_\odot$ black hole¹¹ could be a source of GCR hard radiation but its total luminosity,

evolutionary origin and unique location away from the galactic nucleus would lack a natural explanation. It seems likely that the reported GCR spectrum is a sum of contributions of a few discrete objects located within several degrees of the Galactic-Centre direction; here we explore the possibility that these objects may be low-mass X-ray binaries (LMXBs), a hypothesis suggested by the typical luminosity of accreting neutron stars ($\sim 10^{37}$ erg s⁻¹) and the hard (~ 100 -keV) X-rays with inverse power-law indices $\alpha \approx 2.6$ – 3.0 emitted^{12–15} by some neutron stars.

Arguments and calculations below indicate that a LMXB endowed with an accelerator placed somewhere above the plane of the accretion disk (close to the inner edge of the disk) will give rise to a hard X-ray/soft γ -ray spectrum quite similar to that reported from the GCR. The magnetospheric accretion-powered accelerator serves as a dynamo powering a fireball which shines in hard X-rays on the side of the neutron star and in $\sim$ MeV γ -rays on the side away from the star.

A rapidly rotating magnetized neutron star surrounded by a keplerian accretion disk may be a powerful accelerator of very energetic particles: differences in rotational speed between the star and parts of the disk to which it is coupled by its magnetic field can generate potential drops exceeding 10^{15} V. The maximum power which may be expected¹⁶ from such a dynamo is

$$L_a \approx \frac{(B_* R^3)^2 |\Delta\Omega|}{r_A^3} \approx L_X \frac{R}{r_A} \quad (1)$$

where B_* is the surface dipole field of the neutron star, R is the stellar radius, r_A the radius of the inner edge of the accretion disk, $\Delta\Omega$ is a characteristic difference in angular frequency between the star and relevant parts of the differentially rotating disk, and the X-ray luminosity sustained by accretion is L_X . For the X-ray pulsar Her X-1 with $B_* \approx 4 \times 10^{12}$ G and $\Delta\Omega \approx 4$ s⁻¹, $L_a \approx 10^{35}$ erg s⁻¹. This dynamo has been suggested¹⁷ as the accelerator energy source for the particles which produce the TeV γ -rays claimed to be seen in observations of this binary and others^{18–24}.

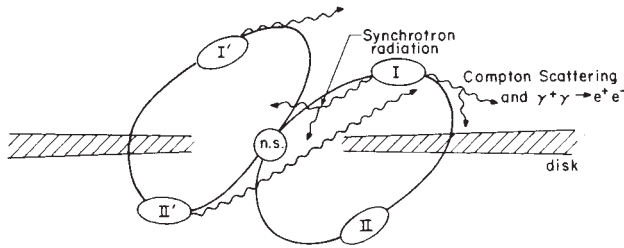


Fig. 1 Schematic geometry of the regions producing proposed hard X-rays and soft γ -rays in some accreting neutron stars (n.s.).

For a near-Eddington-luminosity LMXB the typical values are $B_* \approx 6 \times 10^8$ G, $r_A \approx R$ and $L_{a,38} \equiv L_a / (10^{38} \text{ erg s}^{-1}) \approx 1$. Thus such systems may be even more powerful sources of high-energy particles and non-thermal γ -ray emission than is Her X-1. However, the magnetic field near the interface between the co-rotating magnetosphere of the neutron star and the accretion disk (that is, the region in which particle acceleration, perhaps involving field line reconnection, is expected) is $\sim 10^8$ G in LMXBs, as opposed to $\sim 10^6$ G in Her X-1, whereas the local X-ray flux is >100 times larger in that region for the LMXB system. These two features greatly increase the probability for electron/positron ($e^\pm$) pair production by very energetic photons presumably created in some LMXBs, and also cause subsequent electron synchrotron emission to be much more powerful.

The detailed operation and output of the accelerator is irrelevant in the following as long as the initial energy of the γ -rays or $e^\pm$ it produces greatly exceeds 10^2 MeV and the total power from the accelerator approaches that of equation (1). The LMXB accelerator geometry is assumed to resemble that of Fig. 1. A stream of very-high-energy $e^\pm$ and/or γ -rays is emitted from regions I and II' in the neutron-star direction along the local magnetic field lines. A similar stream moving in the opposite direction (towards the disk) is emitted from the other end of the accelerator. Such streams may (or may not) also be emitted from regions I' and II. Each stream will be the source of an $e^\pm/\gamma$ -ray cascade, in which pair creation by γ -rays on the strong flux of order-keV accretion-powered X-rays of the LMXB plays a crucial role. Some of these accretion X-rays are, in turn, converted into γ -rays by inverse Compton scattering on energetic-shower $e^\pm$ pairs. These processes continue to distribute the total power among the rapidly multiplying $e^\pm$ s and γ -rays until the energy of the γ -rays drops below the threshold of pair creation—this may occur at different energies for the two streams. For the stream flowing towards the star, the threshold occurs at ~ 100 MeV and photons of this energy can travel a long way before creating a pair in the stellar magnetosphere. Synchrotron radiation will be the dominant $e^\pm$ energy-loss mechanism of pairs thus created and a large fraction of energy will then be emitted in hard X-rays, giving a power-law spectrum ($\alpha \approx 3$).

The shower that propagates away from the star does not move into magnetic fields large enough for synchrotron radiation to dominate. Inverse Compton scattering and subsequent pair production by the resulting γ -rays can continue as long as there is a sufficient local flux of hard enough X-rays; under certain conditions an optically thick cloud of order-MeV photons and $e^\pm$ pairs may form. This cloud would be responsible for the MeV bump reportedly seen in the GCR spectrum.

For LMXBs, the magnetic field strength in the region of the gap is $B = B_* \times 10^8$ G with $B_* \approx 1$. The presumed accelerator is at a distance $r = r_* \times 10^6$ cm away from the neutron star, with $1 \leq r_* \leq 10$. The accelerator power satisfies $L_a \approx L_X$, and $L_X \equiv L_{X,38} \times 10^{38} \text{ erg s}^{-1}$ (with $L_{X,38} \approx 0.1$ –1), where L_X is the luminosity of soft X-ray photons of energy $E_X \approx 3 \text{ keV} (L_{X,38})^{1/4}$, typical of the assumed isotropic black-body release of accretion energy. TeV γ -rays will create pairs in the 10^8 G field, and initiate an electromagnetic cascade when the angle between the magnetic field and the photon momentum exceeds $\sim 3 \times 10^{-3}$. Photons of

energy $E_\gamma \geq 40 \text{ GeV}$ create pairs on the magnetic field ($\gamma \rightarrow e^\pm$) or, for lower values of E_γ , create pairs in collision with the soft X-ray photons of energy E_X ; the optical depth (above threshold) for the latter process is $\tau_1 \approx 4 \times 10^3 r_*^{-1} (E_\gamma/\text{GeV})^{-1} L_{X,38}^{1/2}$. The electrons quickly lose their energy to a few photons either through synchrotron radiation, through pair creation on soft X-rays or through inverse Compton scattering, and the process is repeated for the energetic photons thus created until their energy is smaller than the threshold value for pair creation, E_1 . Note that the cascade can be initiated by photons or electrons of energy as low as 1 GeV.

On the side of the neutron star, $E_1 = (m_e c^2)^2 / E_X \approx 100 \text{ MeV} \times L_{X,38}^{-1/4}$ and fireball photons of energy slightly exceeding E_1 can either leave the system (as will all photons of energy $E < E_1$ unless they are intercepted by the disk or the star) or travel a considerable distance before they are absorbed in a collision with an X-ray/ γ -ray photon, thus creating 'last generation' electron pairs of energy, in units of electron rest mass, $\gamma_1 \approx E_1 / (2m_e c^2) \approx 10^2$. Note that as E approaches E_1 from above, the pitch angle ψ of the created pair increases. These electrons will gyrate about the magnetic field lines and each of them will give rise to a (time-integrated) synchrotron spectrum of power-law index $-3/2$. The high-energy cutoff of such an individual spectrum depends on the local value of the magnetic field. The envelope of these spectra (summed over electrons, that is, effectively over the strength of the spatially varying magnetic field) yields a broken power-law photon number spectrum:

$$N(E) \propto \begin{cases} E^{-1.5} & \text{if } E_{\min} < E < E_1 \\ E^{-\beta} & \text{if } E_1 < E < E_2 \end{cases} \quad (2)$$

where $E_{\min} \approx 1 \text{ eV} \times B_8$ (1 eV is the cyclotron energy in a magnetic field of 10^8 G), $E_1 = 10 \text{ keV} (\gamma_1/100)^2 (B_8^\pm)_{\min}$, $E_2 = 10 \text{ keV} (\gamma_1/100)^2 (B_8^\pm)_{\max}$, and $(B_8^\pm)_{\min}$ and $(B_8^\pm)_{\max}$ are the minimum and maximum values of the transverse magnetic field (that is, $B^\perp = B \sin \psi$) encountered by the last-generation pairs. Recall that $(\gamma_1/100) \approx L_{X,38}^{-1/4}$ and thus $E_1 \approx B_* L_{X,38}^{-1/2}$. The shape of the spectrum is sensitive to the spatial distribution of the pair creation rate $R_+(\mathbf{r})$. A dipole field yields $\beta = 2.7$, 3.0 and 3.3, respectively, for $R_+(\mathbf{r}) \propto r^{-1}$, $R_+(\mathbf{r}) = \text{constant}$ and R_+ constant along magnetic flux tubes. Up to a quarter of the dynamo power, that is, $\frac{1}{4} L_a$, is converted to last-generation ($\sim 50 \text{ MeV}$) $e^\pm$ pairs (produced at a rate of $R_i \approx 10^{42} \text{ s}^{-1} r_*^{-2} L_{a,38} \text{ erg s}^{-1}$) and hence to the synchrotron emission described.

The same processes would occur for the outwardly propagating shower (on 'the far side' of the accelerator) if the inequality $B_*^2 \sin^2 \psi > L_{X,38} r_*^{-2}$ were satisfied everywhere. However, because $B^2 \propto r^{-6}$, where r is the distance from the centre of the neutron star, it is likely that at least in some LMXB systems the inequality is violated on the far side, in which case the dominant emission mechanism for the electrons is inverse Compton scattering. Each electron would thus give rise to a photon spectrum of index $\alpha = 1.5$, that is, $N(E) \propto E^{-3/2}$, just as in the synchrotron case, except that now $E_X < E < \gamma_1^2 E_X$. As $L_X \approx L_a$, the number of photons at the high-energy end of this nearly flat spectrum is sufficient to be optically thick to pair creation by photons of energy exceeding $E_b \approx 2(m_e c^2)^2 / (\gamma_1^2 E_X) \approx 1 \text{ MeV}$. The optical depth for this process is $\tau_2 \approx (L_{a,38}) \gamma_1^{-1} \tau_1 \approx 40 (L_{a,38}) r_*^{-1}$. If $\tau_2 > 1$, no photons of energy exceeding E_b will break out on the far side of the accelerator-produced fireball as long as most of the Compton-boosted 'target' X-ray photons are not moving in the same direction as the γ -rays with $E > E_b$. The extent to which these photons with $E > E_b$ will be emitted depends on the field line geometry (possibly quite complicated if magnetic field reconnection is occurring) and on the extent to which target photons from the other showers, including those on the other side of the disk, can be encountered by the γ -rays. If inverse Compton emission dominates on the far side, and $\tau_2 > 1$, the remaining half of the dynamo power is distributed evenly among order-1-MeV photons—which, if observed, would give rise to a

1-MeV bump in the spectrum—and order-1-MeV electrons and positrons. The latter are created at a rate $\frac{1}{4} L_a / 1 \text{ MeV} \approx 10^{43} \text{ s}^{-1} L_{a,38}$. Not knowing the detailed geometry of the accelerator and hence of the fireball or the disk thickness we are unable to predict in what fraction of LMXBs emitting hard X-rays the optically thick cloud of 1-MeV positrons and photons will be formed or the likelihood of MeV-bump observation. We are uncertain about the ultimate fate of the positrons created in this cloud, but if they were channelled along magnetic field lines to the companion star or if they managed to escape the binary system altogether they could contribute to the narrow 511-keV spectrum observed in the GCR spectrum. This line has been observed at a flux corresponding to a pair creation rate of $\sim 5 \times 10^{42} \text{ s}^{-1} \times (d/10 \text{ kpc})^2$ for an isotropic source a distance d away from us, the exact value of the rate depending on the mode of annihilation (the positronium fraction). Perhaps it is significant that copious production of positrons is associated in our model with the MeV bump. Such a correlation has in fact been reported by the HEAO-3 group², but is controversial^{5,25}.

Thus we find that equipping a canonical LMXB with a magnetospheric electron-accelerating mechanism of plausible characteristics leads to the emission of hard X-rays with an inverse power-law spectrum of photon index $\alpha \approx 3$. The power in the hard X-rays is comparable with the total luminosity of the source. At lower energies the power-law index is -1.5 . The break in the spectrum occurs at a value (E_1 in equation (2)) of typically $\sim 10 \text{ keV}$; for a source with lower accretion rates and stronger magnetic field the break may occur at $\sim 100 \text{ keV}$. This could explain the observed hard X-ray emission of such objects as Cyg X-2 (ref. 13), Cyg X-3 (ref. 26), GX5-1 (ref. 15), the pulsating source 1626-67 (ref. 27) and even perhaps 1E1740.7-2942 (refs 9, 10). This mechanism provides the justification for identifying the power-law component of the Galactic-Centre spectrum with the total X-ray emission of a few discrete sources¹⁵. The same mechanism should be operative in strongly magnetized ($B \approx 10^{12} \text{ G}$) accreting neutron stars, for example, GX1+4, but the ratio of hard X-ray luminosity to the total luminosity of these sources would be only $\sim 1\%$. Our model may be applicable to low-luminosity bursting X-ray sources^{28,29}. We predict that accreting neutron stars that emit hard X-rays may be sources of $\sim 10^2$ -MeV γ -rays of comparable luminosity.

Some of the sources may also be copious producers of positrons and order-MeV γ -rays, but this conclusion depends sensitively on the specific parameters of the system, in particular on the unknown geometrical placement of the accelerator. The MeV bump could then be explained by the emission of one or two LMXBs and in principle enough positrons would be produced at the same time to account for the strength of the observed 511-keV annihilation line. It is not clear whether the positrons can diffuse to a suitable annihilation site. We predict that the MeV bump is not a phenomenon unique to the Galactic-Centre direction and urge that the soft γ -ray spectra of relevant accreting neutron stars be carefully observed, particularly for evidence of broad and narrow emission features.

We acknowledge partial support by NASA and NSF grants.

Received 15 April; accepted 26 October 1988.

- Riegler, G. R., Ling, J. C., Mahoney, W. A., Wheaton, W. A. & Jacobson, A. S. in *Positron-Electron Pairs in Astrophysics* (eds Burns, M. L., Harding, A. K. & Ramaty, R.) 230-236 (AIP, New York, 1983).
- Riegler, G. R., Ling, J. C., Mahoney, W. A., Wheaton, W. A. & Jacobson, A. S., *Astrophys. J.* **294**, L13-L15 (1985).
- Riegler, G. R. *et al. Astrophys. J.* **248**, L13-L16 (1981).
- Leventhal, M., MacCallum, C. J., Hutters, A. F. & Stang, P. D. *Astrophys. J.* **260**, L1-L5 (1982).
- Share, G. H. *et al. Astrophys. J.* **326**, 717-732 (1988).
- Blandford, R. in *The Galactic Center* (eds Riegler, G. R. & Blandford, R. D.) 177-179 (AIP, New York, 1982).
- Burns, M. L. in *Positron-Electron Pairs in Astrophysics* (eds Burns, M. L., Harding, A. K. & Ramaty, R.) 281-286 (AIP, New York, 1983).
- Kardashev, N. S., Novikov, I. D., Polnarev, A. G. & Stern, B. E. in *Positron-Electron Pairs in Astrophysics* (eds Burns, M. L., Harding, A. K. & Ramaty, R.) 253-256 (AIP, New York, 1983).
- Skinner, G. K. *et al. Nature* **330**, 544-547 (1987).
- Cook, W. R. *et al. IAU Symp.* **136** (in the press).
- Lingenfelter, R. E. & Ramaty, R. in *The Galactic Center* (eds Riegler, G. R. & Blandford, R. D.) 148-159 (AIP, New York, 1982).

- Dolan, J. F. *et al. Astrophys. J.* **238**, 238-243 (1980).
- Maurer, G. S., Johnson, W. N., Kurfess, J. D. & Strickman, M. S. *Astrophys. J.* **254**, 271-278 (1982).
- Levine, A. M. *et al. Astrophys. J. Suppl.* **54**, 581-617 (1984).
- Knight, F. K., Johnson, W. N., Kurfess, J. D. & Strickman, M. S. *Astrophys. J.* **290**, 557-567 (1985).
- Ruderman, M., Shaham, J., Tavani, M. & Eichler, D. *Astrophys. J.* (in the press).
- Ruderman, M. in *Proc. NATO ASI High Energy Phenomena around Collapsed Stars* (ed. Pacini, F.) 145-191 (Reidel, Norwell, Mass., 1986).
- Dowthwaite, J. C. *et al. Nature* **309**, 691-693 (1984).
- Baltrusaitis, R. M. *et al. Astrophys. J.* **293**, L69-L72 (1985).
- Gorham, P. W. *et al. Astrophys. J.* **309**, 114-121 (1986).
- Chadwick, P. M. *et al. in Very High Energy Gamma Ray Astronomy* (ed. Turver, K. E.) 121-127 (Reidel, Dordrecht, 1987).
- Lamb, R. C. in *13th Texas Symposium on Relativistic Astrophysics* (ed. Ulmer, M. P.) 589-594 (World Scientific, Singapore, 1987).
- Lamb, R. C. *et al. Astrophys. J.* **328**, L13-L16 (1988).
- Resvanis, L. K. *et al. Astrophys. J.* **328**, L9-L12 (1988).
- Leventhal, M., MacCallum, C. J. & Stang, P. D. *Astrophys. J.* **225**, L11-L14 (1982).
- White, N. E. & Holt, S. S. *Astrophys. J.* **257**, 318-337 (1982).
- Pravdo, S. H. *et al. Astrophys. J.* **231**, 912-918 (1979).
- White, N. E. & Mason, K. O. *Space Sci. Rev.* **40**, 167-194 (1985).
- White, N. E., Stella, L. & Parmar, A. N. *Astrophys. J.* **324**, 363-378 (1987).

Constraints on the equation of state for neutron stars implied by PSR1957+20

John L. Friedman*, James N. Imamura†, Richard H. Durisen‡ & Leonard Parker*

* Department of Physics, University of Wisconsin-Milwaukee, Wisconsin 53201, USA

† Institute of Theoretical Science and Department of Physics, University of Oregon, Eugene, Oregon 97403, USA

‡ Department of Astronomy, Indiana University, Bloomington, Indiana 47405, USA

The primary uncertainty in understanding the structure of neutron stars is our lack of knowledge of the equation of state of condensed matter above nuclear density ($3 \times 10^{14} \text{ g cm}^{-3}$). The most rapidly rotating neutron stars, the millisecond pulsars, may offer a way to probe the equation of state in this region. The periods of the two fastest known pulsars, PSR1937+214 and the recently discovered¹ PSR1957+20, are within 3% of each other, and are shorter than the period of any other known pulsars by about a factor of two. This observation substantially strengthens the argument that the fastest pulsars should rotate at the limiting frequency for stable neutron stars^{2,3}. We show here that if the two fastest pulsars are rotating at or near their limiting frequencies, and if their masses are at least $1.4 M_\odot$, then the equation of state of neutron star matter must be quite stiff.

The frequencies of PSR1957+20 and PSR1937+214 are 4,033 and 3,910 radians s^{-1} , respectively. Of 11 equations of state used in the tabulated models of Arnett and Bowers⁴, only models M (tensor interaction, Pandharipande and Smith⁵) and L (mean field, Pandharipande and Smith⁵) are stiff enough to allow a limiting frequency, Ω_{lim} , as small as 3,910 radian s^{-1} for masses $M > 1.4 M_\odot$. (A stiff equation of state has a larger pressure, at a given density, than a soft equation of state.) Each of these pulsars is spinning down, and so must now be rotating at least slightly slower than Ω_{lim} . However, if the close agreement in frequency is not coincidental, but reflects the fact that each frequency is within a few per cent of its gravitational limit, the same two equations of state are picked out. If one assumes only that the frequency of each pulsar is within 25% of its limiting value, only the four stiffest equations of state are consistent with $M > 1.4 M_\odot$.

The limiting frequency, Ω_{lim} , decreases with increasing stiffness in the equation of state, because, for a given mass, models of neutron stars constructed from equations of state that are stiff above nuclear density have substantially larger radii and moments of inertia than models based on the softer equations of state. A larger radius implies a smaller value of Ω_{lim} . In particular, a firm upper limit on the angular velocity of uniformly rotating stars is the Kepler frequency Ω_K , the

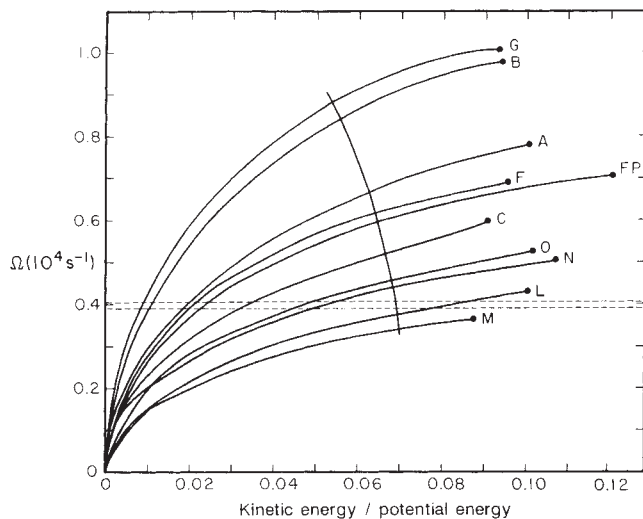


Fig. 1 Sequences of uniformly rotating neutron stars of baryon mass $M_B = 1.4 M_\odot$ for equations of state tabulated by Arnett and Bowers⁴ (indicated by letters), together with the more recent Friedman-Pandharipande²⁵ equation of state (FP). The Keplerian termination point of each sequence is indicated by a dot. The diagonal line crossing the sequences is a conservative estimate (that is, an underestimate) of the smallest rotation for which the models could be unstable to non-axisymmetric perturbations. The Ω values for PSR1937+214 and PSR1957+20 are indicated by the upper and lower dashed lines, respectively. For a given equation of state and mass, the limiting frequency almost certainly lies in the narrow band of frequencies between the diagonal line and the termination point. Because the curves for larger masses lie above these, if PSR1937+214 and PSR1957+20 each have $M_B > 1.4 M_\odot$ and rotate at their limiting frequencies, the equation of state must be at least as stiff as L.

frequency of a particle in circular orbit at the star's equator. Instability to non-axisymmetric perturbations, driven by gravitational radiation, may well imply that Ω_{lim} is smaller than Ω_K , but the value of the angular velocity at the instability point differs only slightly (by less than 10%) from Ω_K (see ref. 6 and the discussion below). In contrast, for the various proposed equations of state, Ω_{lim} varies from 3,500 to 10,000 radian s^{-1} for neutron stars of baryon mass $1.4 M_\odot$. Thus, observation of the precise value of Ω_{lim} stringently constrains the equation of state for neutron stars, if their masses are known⁶⁻⁸.

The non-axisymmetric instability of rotating relativistic stars is driven by gravitational radiation, and has a growth time τ_{gr} determined by the rate at which energy is radiated. The relevant modes have an angular dependence $\exp(im\phi)$, where ϕ is the azimuthal coordinate⁹. In the absence of viscosity and thermal conductivity, the shortest-wavelength modes (large m) become unstable at the lowest values of Ω . But the short-wavelength modes also have the longest growth time, and, when viscosity is present, a mode is stable if the viscous damping time is shorter than the growth time τ_{gr} . In neutron stars, the viscosity (ν) rises rapidly with decreasing temperature ($\nu \propto T^{-2}$) and for $T < 10^6$ K all modes are likely to be stable. For $T > 10^8$ K, the non-axisymmetric instability can be expected to set the upper limit on rotational frequency^{3,7,10-11}.

An expected temperature of 10^7 K for accreting neutron stars leaves open the question of whether the frequency is limited by the non-axisymmetric instability or simply by Ω_K , but the constraint on the equation of state is not sensitive to the limiting mechanism. Friedman, Ipser and Parker⁶ have calculated the value of Ω_K for a wide range of neutron-star masses and equations of state. In terms of the dimensionless measure of rotation, $t = K/|W|$ (where K is the bulk kinetic energy and W is the gravitational potential energy), all sequences of uniformly rotating neutron stars terminate at $t < 0.13$, which is much smaller

than the value of 0.5 allowed by differential rotation (see Fig. 1). On the other hand, a determination of the instability points for newtonian models by Imamura *et al.*¹² and Managan¹³ requires that t be at least as large as that corresponding to the diagonal line in Fig. 1. The diagonal line is an approximation to the $m = 4$ instability point with zero viscosity, using the values of $K/|W|$ obtained from newtonian stars with a polytropic index comparable to an average value for the equation of state of the plotted models. One expects a kinematic viscosity ν greater than $1 \text{ cm}^2 \text{ g}^{-1}$, and if the $m = 3, 4$ and 5 modes become unstable, the instability points then lie to the right of this line. The viscosity is almost certainly large enough to make modes with $m > 5$ stable for the entire equilibrium sequences. A nearly identical line is obtained from Lindblom's independent estimate of the onset of instability in the presence of viscosity¹¹. A later suggestion by Lindblom¹⁴ that viscosity may lead to instability of an $l = m = -2$ mode just before the point at which $\Omega = \Omega_K$ seems unlikely, because all sequences terminate before the value $t = 0.14$ at which $m = \pm 2$ modes become unstable for newtonian stars. Because of the flat slope of the plot of Ω against t , the value of Ω_{lim} is narrowly constrained at a given mass for each equation of state, regardless of the limiting mechanism. Also, for fixed equation of state, the limiting frequency of a neutron star increases with its mass. Thus, if one assumes baryon masses at least as large as $1.4 M_\odot$ (or gravitational masses greater than $1.3 M_\odot$), Fig. 1 shows that PSR1937+214 and PSR1957+20 can be rotating at their limiting frequency only if the equation of state is as stiff as those of models L or M (ref. 5).

There are now seven known pulsars with rotation periods < 10 ms, of which only two are not in binary systems^{1,15-19}. As a class these millisecond pulsars have small values of period derivative $\dot{P}$, in the sense that their spin-down rates imply large pulsar ages and small spin-down magnetic field strengths²⁰. They are thought to be either neutron stars spun-up by accretion in low-mass X-ray binaries (LMXB)²¹ or to be the remnants of accretion-induced collapses of white dwarfs in cataclysmic variable systems²². Both scenarios appear to be capable of explaining the general properties of millisecond pulsars. In the LMXB scenario, a magnetic neutron star which is initially slowly rotating is spun-up by the accretion of angular-momentum-rich material from its companion star. If the accretion disk extends all the way down to the star's equator, a $1.4 M_\odot$ neutron star can be spun-up to millisecond periods by accreting $\sim 0.2 M_\odot$ of material. The precise result is sensitive to the star's moment of inertia, and hence to the equation of state. Using the results of Friedman, Ipser and Parker⁶, we find that for the equations of state of model N an initially non-rotating $1.4 M_\odot$ neutron star can be spun-up to a period of 1.558 ms if it accretes a mass of $0.18 M_\odot$, whereas for equation of state L it must accrete $0.25 M_\odot$. Only a small fractional mass accretion is needed to achieve Ω_{lim} . On the other hand, the difference of 0.1 ms^{-1} in limiting frequency between PSR1937+214 and PSR1957+20 corresponds to a mass difference of the order of $0.1 M_\odot$, arising either from a difference in initial mass or a difference in additional mass accreted once the limiting frequency was reached.

If the magnetic field of the neutron star is sufficient to hold the inner edge of the accretion disk away from the equator of the star, the minimum rotation period is that at which the spin-up torque arising from the accreting matter balances the stress from the interaction of the accretion disk and the magnetic field of the neutron star²¹. This occurs when the rotation of the neutron star is nearly synchronous with the inner edge of the accretion disk. The equilibrium spin period is approximately given by

$$P_{\text{eq}} = 1.9 (B/10^9 \text{ G})^{6/7} (M/M_\odot)^{-5/7} \times (\dot{M}/\dot{M}_E)^{-3/7} (R/10 \text{ km})^{18/7} \text{ ms} \quad (1)$$

where B is the magnetic field strength, $\dot{M}$ is the accretion rate, R is the star's radius and $\dot{M}_E$ is the Eddington accretion rate²¹. It is conceivable that this process limits the rotation rates of

most millisecond pulsars (see Fig. 2 in ref. 23). It requires fine-tuning of B , M and $\dot{M}$, however, to make the two fastest pulsars have nearly the same periods. In any case, given that PSR1821-24 lies to the left of the spin-up limit defined by equation (1) in the $(P, \dot{P})$ plane¹⁸, this equation may be valid only to an order of magnitude. We find it unlikely that a many-parameter mechanism such as that described in equation (1) would lead to the well-defined observed upper limit on the rotation rates of millisecond pulsars. Because the gravitationally limited frequency Ω_{lim} depends only on the mass of the neutron star, we consider such a mechanism to be much more convincing in explaining the apparent existence of a precise maximum frequency for the millisecond pulsars.

There has been a recent suggestion²⁴ that in old neutron-star systems, where the magnetic fields have decayed to 10^8 – 10^9 G, the magnetic torque mechanism described above depends on fewer parameters than implied by equation (1). If this is true, then the magnetic torque mechanism might be able to account for a sharp cutoff in the period distribution of the millisecond pulsars. But the magnetic fields of the two fastest pulsars, inferred from their observed spin-down rates ($\dot{P} = 1.05 \times 10^{-19} \text{ s s}^{-1}$ for PSR 1937+214 and $\dot{P} < 3 \times 10^{-20} \text{ s s}^{-1}$ (A. S. Fruchter, D. R. Stinebring and J. H. Taylor, personal communication) for PSR1957+20), differ by a factor of at least 1.8. This appears to be too large to be consistent with a 3% agreement in observed frequencies if this arises from a magnetically determined upper limit.

In the white-dwarf-collapse scenario, a millisecond pulsar is formed if the rotation period of the white dwarf is $< 1,000$ s. The fastest observed white-dwarf rotator, AE Aqr, has a rotation period of 33 s. Thus, if neutron stars with sufficiently weak magnetic fields can form from the collapse of white dwarfs, one would expect their frequencies to range up to Ω_{lim} . Neutron stars with $\Omega = \Omega_{\text{lim}}$ could arise from white dwarfs that initially rotate slightly too fast to form stable neutron stars and which are able to rid themselves of their excess angular momentum

during collapse, by shedding matter or radiation (for example, by gravitational radiation from an initially unstable neutron star). Again, in this scenario, the existence of a well-defined upper limit on millisecond pulsar rotation frequencies suggests that $\Omega = \Omega_{\text{lim}}$ for the fastest rotators.

We conclude that the close correspondence of the periods for the two fastest pulsars known suggests that their rotation periods are limited by break-up or by a gravitational-radiation-reaction-driven instability, that the equation of state for neutron star matter is stiff, and that the masses of these two pulsars differ by only a few per cent. These conclusions would be strengthened by the discovery of additional pulsars with $P \approx 1.5$ – 1.6 ms and by the continued absence of pulsars with shorter periods.

This work was supported by the NSF.

Received 7 July; accepted 21 October 1988

- Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. *Nature* **333**, 237–238 (1988).
- Imamura, J. N., Middleditch, J. & Steiman-Cameron, T. Y. *Astrophys. J.* **314**, L11–L13 (1987).
- Friedman, J. L. *Phys. Rev. Lett.* **51**, 11–14 (1983).
- Arnett, W. D. & Bowers, R. L. *Astrophys. J. Suppl.* **33**, 415–436 (1977).
- Pandharipande, V. & Smith, R. A. *Nucl. Phys. A* **175**, 225 (1975).
- Friedman, J. L., Ipser, J. R. & Parker, L. *Astrophys. J.* **304**, 115–139 (1986).
- Lindblom, L. *Astrophys. J.* **303**, 146–153 (1986).
- Managan, R. A. *Astrophys. J.* **309**, 598–608 (1986).
- Friedman, J. L. & Schutz, B. R. *Astrophys. J.* **222**, 281–296 (1978).
- Wagoner, R. V. *Astrophys. J.* **278**, 345–348 (1984).
- Cutler, C. & Lindblom, L. *Astrophys. J.* **314**, 234–241 (1987).
- Imamura, J. N., Friedman, J. L. & Durisen, R. H. *Astrophys. J.* **294**, 474–478 (1985).
- Managan, R. A. *Astrophys. J.* **294**, 463–473 (1985).
- Lindblom, L. *Astrophys. J.* **317**, 325–332 (1987).
- Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M. & Goss, W. M. *Nature* **300**, 615–618 (1982).
- Boriakoff, V., Buccheri, R. & Fauci, F. *Nature* **304**, 417–419 (1983).
- Segelstein, D. J., Rawley, L. A., Stinebring, D. R., Fruchter, A. S. & Taylor, J. H. *Nature* **322**, 714–717 (1986).
- Foster, R. S., Backer, D. C., Taylor, J. H. & Goss, W. M. *Astrophys. J.* **326**, L13–L16 (1988).
- Ables, J. G. & Jacka, C. E. *IAU Circ.* no. 4602 (1988).
- Taylor, J. H. & Stinebring, D. R. *Astr. Astrophys.* **24**, 285–327 (1986).
- Alpar, M. A., Cheng, A. F., Ruderman, M. A. & Shaham, J. *Nature*, **300**, 728–730 (1982).
- Chanmugam, G. & Brechet, K. *Nature* **329**, 696–697 (1987).
- Dewey, R. J., Maguire, C. M., Rawley, L. A., Stokes, G. H. & Taylor, J. H. *Nature* **322**, 712–714 (1986).
- Wasserman, I. & Cordes, J. *Astrophys. J.* **333**, L91–L94 (1988).
- Friedman, B. & Pandharipande, V. R. *Nucl. Phys. A* **361**, 502 (1981).

Measurements of H_2O_2 and SO_2 in clouds and estimates of their reaction rate

A. S. Chandler*, T. W. Choularton*, G. J. Dollard†, A. E. J. Eggleton†, M. J. Gay*, T. A. Hill*, B. M. R. Jones†, B. J. Tyler‡, B. J. Bandy§ & S. A. Penkett§

* Department of Physics, University of Manchester Institute of Science and Technology, Manchester M60 1QD, UK

† Environmental and Medical Sciences Division, The Harwell Laboratory, Didcot, Oxfordshire OX11 0RA, UK

‡ Department of Chemistry, University of Manchester Institute of Science and Technology, Manchester M60 1QD, UK

§ School of Environmental Sciences, University of East Anglia, Norwich, Norfolk NR4 7TJ, UK

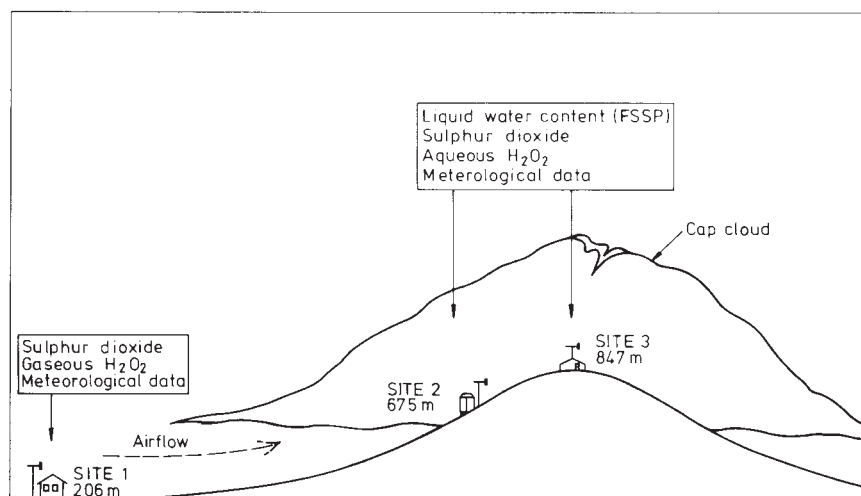
Recently the oxidation of sulphur dioxide (SO_2) in the atmosphere to form sulphuric acid has become an important research area related to the formation of acid rain and subsequent damage to sensitive ecosystems. Mechanisms leading to the formation of sulphuric acid include both homogeneous gas-phase reactions and multi-phase processes occurring in cloud and rain water¹. Laboratory and theoretical studies have demonstrated the importance of hydrogen peroxide (H_2O_2) in oxidizing SO_2 to sulphate^{2–4}. Recent advances in analytical techniques^{5,6} have enabled data to be collected under ambient conditions. These show that an anti-correlation often exists between the simultaneous gas-phase SO_2 and the liquid-phase H_2O_2 concentrations, suggesting that a reaction occurs between SO_2 and H_2O_2 (refs 7–9). Here we report measurements that allow estimation of a field-derived rate constant for

the reaction of SO_2 with H_2O_2 in cloud water. These estimates are significantly higher than those derived from laboratory experiments.

All the experimental work reported here was carried out on or near to Great Dun Fell (GDF) in the north of England. GDF is 847 m above sea level and forms part of the long ridge of the Pennine Hills, which run south-easterly to the north-east of the Eden Valley. The cloud dynamics and microphysics of this site have been studied extensively¹⁰, making GDF a most suitable location for cloud chemistry studies.

The site layout is depicted in Fig. 1; three sites are operated which essentially form a line running south-west to north-east up the fellside. The base site (site 1) is 10 km upwind of GDF, in the Eden Valley and at ~206 m above sea level. Here the air stream entering the cloud is monitored for SO_2 and, occasionally, for gas-phase H_2O_2 . Site 2 is situated on the front face of GDF, ~675 m above sea level, and site 3 is at the summit. The distance between sites 2 and 3 is ~0.5 km. At both of the GDF sites gaseous sulphur dioxide and ozone are measured, together with the cloud liquid-water content and droplet size distribution, using Knollenberg Forward Scattering Spectrometer Probes. The SO_2 measurements were made with a temperature-thermostated flame-photometric detector (SAE 285, Meloy Instruments), which had a detection limit of ~0.2 parts per 10^9 by volume (p.p.b.v.) of SO_2 . Zero checks were made every 10 to 15 minutes and calibrations were performed every 30 minutes. A second, fluorimetric SO_2 analyser was operated at site 3, the results from which agreed closely with those from the former instrument. The gas-phase H_2O_2 measurements at site 1 were made with an instrument that used aqueous collection of H_2O_2 from air followed by its determination using an enzyme fluorescence technique¹¹. Passive cloudwater collectors⁷ were deployed at each

Fig. 1 Experimental layout at Great Dun Fell, showing the relative positioning of the three observation sites.

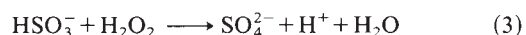
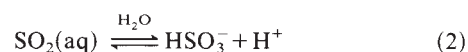
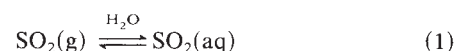


site for continuous collection of cloud water, in which hydrogen peroxide was monitored using a luminol chemiluminescence in the presence of a microperoxidase catalyst⁵. The analysers used at the different sites were all operated and calibrated with the same batches of reagents and solutions, and had a detection limit of less than 50 nmol l⁻¹. Cloudwater pH was measured simultaneously and the cloud water was stored for subsequent chemical analysis. Standard meteorological measurements of windspeed, wind direction, temperature and humidity were made at all three sampling sites.

Figure 2 is an example of a time-trend trace for SO₂ and H₂O₂, obtained on 19 May 1986. There is clear evidence that hydrogen peroxide in cloud water is consumed during periods when pulses of gas-phase SO₂ occur, that is, as air with elevated background SO₂ concentration advects across the site. The oxidation reaction must occur in the cloud droplets during the transit of air from cloud base to the observation sites (sites 2 and 3). Reaction on aqueous aerosol outside the cloud envelope¹² is not thought to be important in these results, as the H₂O₂ concentration in air entering cloud base was independent of the SO₂ concentration changes. With certain assumptions, therefore, it is possible to determine the rate of SO₂ oxidation by H₂O₂ in the cloud system.

Aqueous-phase oxidation of SO₂ by H₂O₂ in the laboratory has been studied by several workers^{2,13-16}. The relevant

reactions are



From Martin and Damschen¹⁵ the rate of formation of SO₄²⁻ from SO₂ is given by

$$\frac{d[\text{SO}_4^{2-}]}{dt} = k'[\text{H}_2\text{O}_2][\text{SO}_2(\text{aq})] \text{ l mol s}^{-1} \quad (4)$$

where [H₂O₂] is the peroxide concentration in solution and [SO₂(aq)] is related to the gas-phase SO₂ concentration by the Henry's law coefficient¹⁷. Here, *k'* is a rate factor that includes the first ionization constant for sulphurous acid and the equilibria established with intermediate species, after the mechanism of Hoffman and Edwards¹⁴. Equation (4) is a convenient form of the rate equation as it is essentially insensitive to pH in the pH range 1–5, which includes most of the pH values observed in the present study.

A field study of this reaction ideally should relate sulphate production to peroxide decay. But the measurement errors associated with this were large as, in most situations studied

Table 1 Second-order rate factor for SO₂ oxidation by H₂O₂

Date	Time GMT	Observation site	Transit time* <i>T</i> (s)	[SO ₂] (p.p.b.v.)	Initial† [H ₂ O ₂] _i (μmol l ⁻¹)	Final† [H ₂ O ₂] _f (μmol l ⁻¹)	pH	Liquid water content (g m ⁻³)	<i>k'</i> ‡ (l mol ⁻¹ s ⁻¹ × 10 ⁶)
9/5/86	14:00	2	120	1.0	38(0.30)§	2	5.2	0.35	9.4
	15:00	2	127	1.0	38(0.43)§	5	5.6	0.40	6.1
	15:00	2	127	0.2	9	5	5.6	0.40	7.1
	16:00	2	183	1.0	45(0.25)§	11	5.6	0.60	2.9
	14:00	3	170	1.0	24(0.30)§	15	5.2	0.50	6.2
	16:00	3	273	1.0	37(0.25)§	7	5.6	0.60	2.3
14/5/86	15:00	3	150	1.0	112(0.50)§	7	5.4	0.20	6.9
19/5/86	09:20	2	160	0.5	37	8	4.3	0.55	7.3
	09:20	3	460	0.5	31	5	4.3	0.50	3.0
	12:00	3	300	0.5	35	5	4.2	0.25	5.0
	13:00	3	80	1.0	70	10	4.0	0.10	9.3
20/5/86	17:00	2/3	117	1.1	20	5	4.1	0.40	4.1
Average									5.8 ± 2.4

* Time taken for air parcel to move from cloud base to the observation site.

† See text for use.

‡ Derived using $d[\text{SO}_4^{2-}]/dt = k'[\text{H}_2\text{O}_2][\text{SO}_2(\text{aq})]$ (ref. 15).

§ Gas-phase hydrogen peroxide measured in inflow air (p.p.b.v.).

|| pH determined on stored samples.

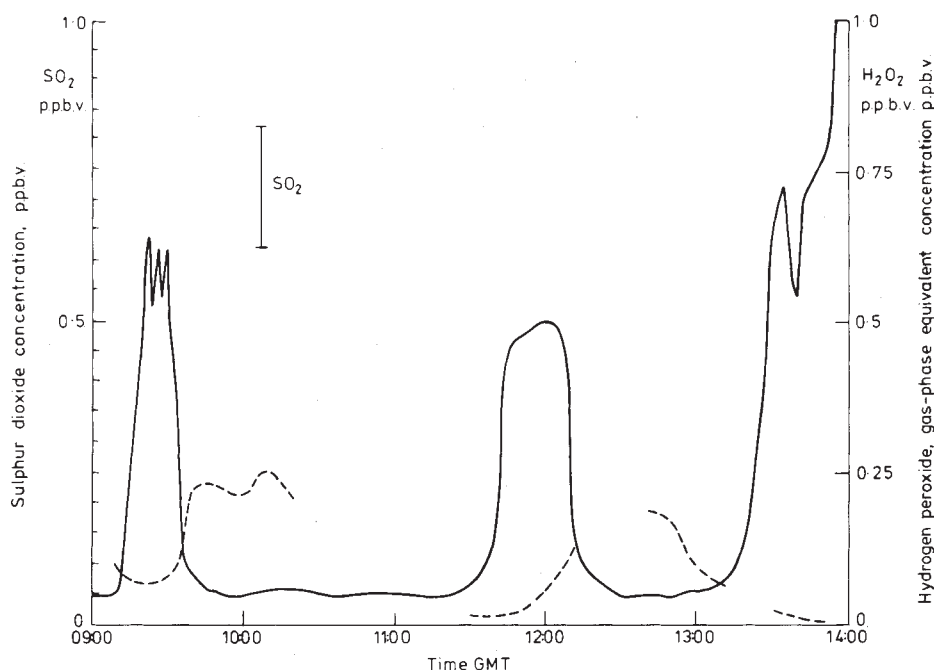


Fig. 2 An example of a time-trend trace for SO_2 and H_2O_2 measured simultaneously in cloud at Great Dun Fell on 19 May 1986. The vertical bar represents the minimum detectable change in the SO_2 signal. The solid line represents SO_2 , and the dashed line represents H_2O_2 .

below, the additional sulphate produced was only a small fraction of the total sulphate present in the cloud water from uptake of incoming aerosol material. The rate data determined was therefore restricted to a study of the peroxide decay. It was observed that H_2O_2 decay occurred only when SO_2 was present in the incoming air. A material balance between sulphate and peroxide has been obtained in related experiments carried out at GDF by other workers¹⁸.

The equilibrium conditions in the cloud are such that most of the hydrogen peroxide enters the liquid phase very rapidly, leaving a relatively low partial pressure in the interstitial air. Thus after a few seconds the liquid phase is effectively a closed system with respect to hydrogen peroxide. Conversely, most of the sulphur dioxide remains in the gas phase, in stoichiometric excess (~3:1). The liquid phase is therefore an open-phase system with respect to the sulphur species. This results in a relatively constant concentration of S(IV) species in solution, as the rate of removal of S(IV) by oxidation is slower than the rate of replenishment from the gas phase.

The decay of H_2O_2 was assumed to follow pseudo-first-order kinetics, and a first-order rate constant was derived for H_2O_2 loss during the transit time (T) of the air mass from cloud base to the observation site. (The H_2O_2 aqueous-phase concentrations all refer to the appropriate observation site; the gas-phase concentrations are in p.p.b.v., measured at site 1.) Thus,

$$\frac{d}{dt}[\text{SO}_2] = -\frac{d}{dt}[\text{H}_2\text{O}_2] = k_1[\text{H}_2\text{O}_2] \quad (5)$$

where

$$k_1 = k'[\text{SO}_2(\text{aq})] \quad (6)$$

and

$$\ln \frac{[\text{H}_2\text{O}_2]_0}{[\text{H}_2\text{O}_2]_t} = k_1 T \quad (7)$$

$[\text{H}_2\text{O}_2]_0$ and $[\text{H}_2\text{O}_2]_t$ are the hydrogen peroxide concentration entering cloud base and that at the observation site, respectively. The transit time up the hill, T , was calculated using the model of Hill *et al.*¹⁹. Measurements made at site 1 showed that the concentration of gas-phase H_2O_2 entering the cloud did not alter substantially during the transit period.

Three methods were applied to calculate k_1 , and thus k' , in equation (4)¹⁵. In the first, the measured concentrations of aqueous H_2O_2 at sites 2 or 3, before the appearance of the SO_2

pulse, were used to define a cloud-base (initial) hydrogen peroxide concentration. In the second method, gas-phase measurements at site 1 were used to calculate the theoretical equivalent H_2O_2 solution-phase concentrations in cloud water at cloud base. In both cases, the observed decline in the hydrogen peroxide concentration, from the measured or theoretical cloud-base value to those values measured at sites 2 and 3 following the SO_2 pulse, was attributed to reaction with SO_2 within the cloud.

For the data gathered on 20 May 1986 a third, slightly different, technique was used. For this day it was possible to use the difference in measured hydrogen peroxide concentrations between sites 2 and 3 to determine the rate constant. This was in fact the original objective of the experiment, but it was found from liquid-water-content data that dry-air entrainment often made the small differences in hydrogen peroxide concentrations recorded at the sites difficult to interpret. In each case the hydrogen peroxide concentrations were corrected for changes in liquid-water content between the sites and for exchange between the gaseous and aqueous phases.

The rate factors (k') derived using these three approaches are summarized in Table 1. When expressed in the pH-independent manner of equation (4), the average value of k' is $5.8 \pm 2.4 \times 10^6 \text{ l mol}^{-1} \text{ s}^{-1}$, which is considerably larger than the value of $8.3 \times 10^5 \text{ l mol}^{-1} \text{ s}^{-1}$ given in ref. 15 for laboratory measurements at 25 °C. As a check on the accuracy of these procedures and the assumptions made above, the model of Hill *et al.*¹⁹ was run using the field-deduced reaction rate for each case study. Using appropriate input conditions, the computed hydrogen peroxide concentrations at the observation sites were within 20% of those observed.

The rate factor k' derived in Table 1 is in fact a composite of the second order rate constant for the reaction of H_2O_2 with HSO_3^- and an equilibrium constant governing the dissociation of $\text{SO}_2(\text{aq})$ into HSO_3^- and H^+ . The value of k' changes little with temperature because of compensating effects on the solubility of SO_2 .

We have demonstrated that the oxidation of SO_2 by H_2O_2 in cloud droplets is extremely rapid and thus accounts for previous observations of the mutually exclusive nature of SO_2 and H_2O_2 in cloud systems⁷⁻⁹. Furthermore, the rate constant determined in the field experiments is substantially larger than that obtained in laboratory studies^{2,14,15}.

The uncertainties associated with field experiments are, of

course, very large, but it is most unlikely that any combination of experimental errors could account for this discrepancy. It is unlikely that the reaction of hydrogen peroxide with S(IV), as observed in our field experiments, is catalysed by weak organic acids (the reaction taking place in cloud at GDF does not appear to be subject to general acid catalysis). Our conclusion is therefore that the oxidation by peroxide of SO₂ in cloud droplets to produce sulphuric acid is extremely rapid.

This work was supported by the Department of the Environment and the NERC.

Received 4 August; accepted 2 November 1988.

1. SO₂, NO and NO₂ Oxidation Mechanisms: Atmospheric considerations (ed. Calvert, J. G.) Acid Precipitation series Vol. 3 (Butterworth, Stoneham, Massachusetts, 1984).
2. Penkett, S. A., Jones, B. M. R., Brice, K. A. & Eggleton, A. E. *J. Atmos. Envir.* **13**, 123-137 (1979).
3. Middleton, P., Kiang, C. S. & Mohnen, V. A. *Atmos. Envir.* **14**, 463-472 (1980).
4. Chameides, W. L. & Davis, D. D. *J. geophys. Res.* **87**, 4863-4877 (1982).
5. Ames, D. L. *Central Electricity Generating Board Rep. no. TPRD/L/2552/N83* CERL Laboratories, Leatherhead, England (1983).
6. Lazrus, A. L., Kok, G. L., Gitlin, S. N., Lind, J. A. & McLaren, S. E. *Anal. Chem.* **57**, 917-922 (1985).
7. Kadlec, J., McLaren, S., Camarota, N., Mohnen, V. & Wilson, J. *Proc. 4th Int. Conf. on Precipitation Scavenging* Santa Monica A Vol. 1 (1983).
8. Daum, P. H., Kelly, T. J., Schwartz, S. E. & Newman, L. *Atmos. Envir.* **18**, 2671-2684 (1984).
9. Kelly, T. J., Daum, P. H. & Schwartz, S. E. *J. geophys. Res.* **90**, 7861-7871 (1985).
10. Choularton, T. W. *et al. Q. J. R. met. Soc.* **112**, 131-148 (1986).
11. Lazrus, A. L. *et al. Anal. Chem.* **58**, 594-597 (1985).
12. Heikes, B. G., Kok, G. L., Walega, J. G. & Lazrus, A. L. *J. geophys. Res.* **92**, 915-931 (1987).
13. Mader, P. M. *J. Am. chem. Soc.* **80**, 2634-2639 (1985).
14. Hoffman, M. R. & Edwards, J. O. *J. phys. Chem.* **79**, 2096-2098 (1975).
15. Martin, L. R. & Damschen, D. E. *Atmos. Envir.* **15**, 1615-1621 (1981).
16. McArdle, J. V. & Hoffmann, M. R. *J. phys. Chem.* **87**, 5425-5429 (1983).
17. National Bureau of Standards *NBS tech. Note* 270-1 124 (1965).
18. Gervat, G. P. *et al. Nature* **333**, 241-243 (1988).
19. Hill, T. A., Choularton, T. W. & Penkett, S. A. *Atmos. Envir.* **20**, 1763-1771 (1986).

Chemically sensitive structure-imaging with a scanning transmission electron microscope

S. J. Pennycook & L. A. Boatner

Solid State Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee 37831-6024, USA

Conventional high-resolution electron microscopy uses the phase-contrast method¹, in which the diffracted beams emerging from the sample are recombined on the viewing screen of the microscope. The resultant contrast depends on the relative phases of the diffracted beams, which is sensitive to microscope and sample parameters, so that images must be interpreted by means of simulation, and defect models are somewhat empirical. By using a high-angle detector in a scanning transmission electron microscope, these problems may be avoided and high atomic-number contrast may be obtained. Here we present results of this technique applied to single crystals of the high-transition-temperature superconductors YBa₂Cu₃O_{7-x} and ErBa₂Cu₃O_{7-x}. The heavy-atom planes are directly imaged as bright lines, and the probable structure of an observed defect is directly inferred from its image.

In the phase-contrast method of high-resolution electron microscopy, introduced by Scherzer¹, the contrast is controlled by the sample thickness and objective lens defocus as well as the atomic positions and species, so that phase-contrast images must be interpreted through image simulation, and defect models are often suggested by intuition or from known defects in structurally similar materials. The high-angle annular detector of a scanning transmission electron microscope (STEM) avoids these structure-sensitive effects by detecting the electrons scattered, predominantly elastically, through high angles. The (incoherent) scattering at such small impact parameters occurs from largely unscreened nuclei, and therefore gives an image showing strong atomic-number (*Z*) contrast with a minimum dependence on specimen thickness and microscope defocus. Such a detector

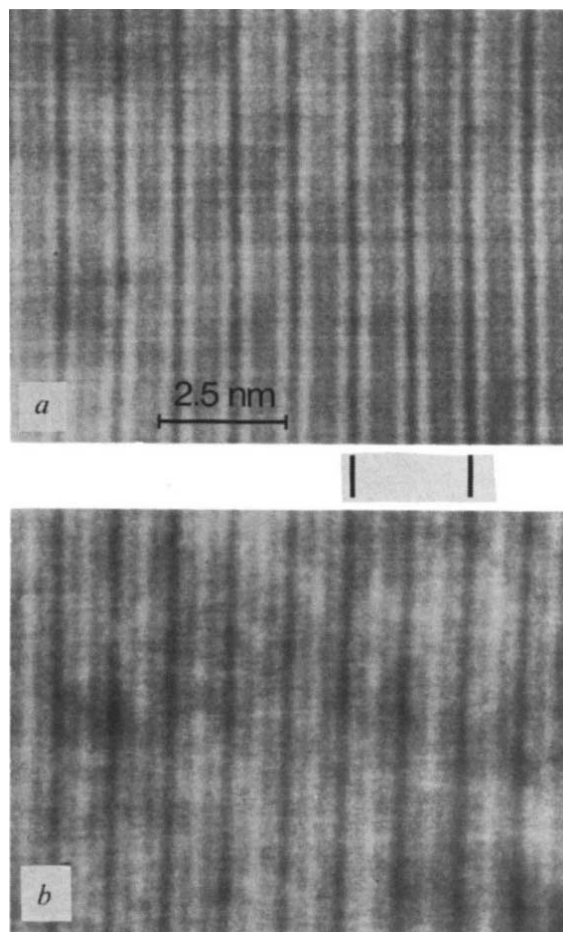


Fig. 1 Z-contrast STEM images of *a*, YBa₂Cu₃O_{7-x} and *b*, ErBa₂Cu₃O_{7-x} single crystals viewed parallel to the *c* planes (probe size ~0.25 nm, incident beam convergence semi-angle = 6 mrad, detector semi-angles = 50-150 mrad).

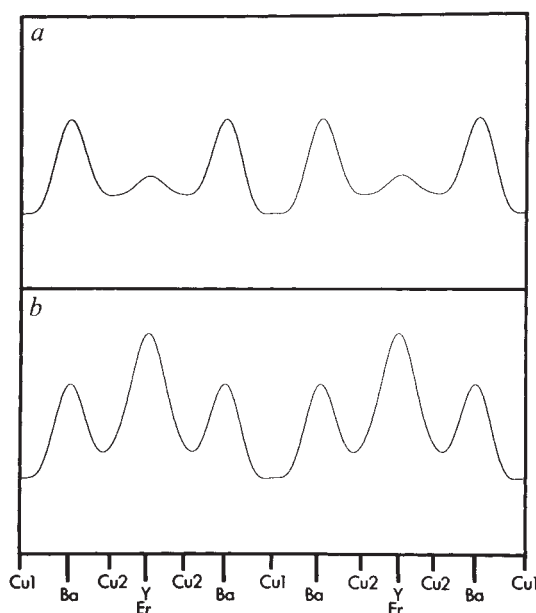


Fig. 2 Calculated image intensity across two unit cells of *a*, YBa₂Cu₃O_{7-x} and *b*, ErBa₂Cu₃O_{7-x}, assuming no electron-channelling effects. Traces should be compared to the images in Fig. 1 between the vertical markers.

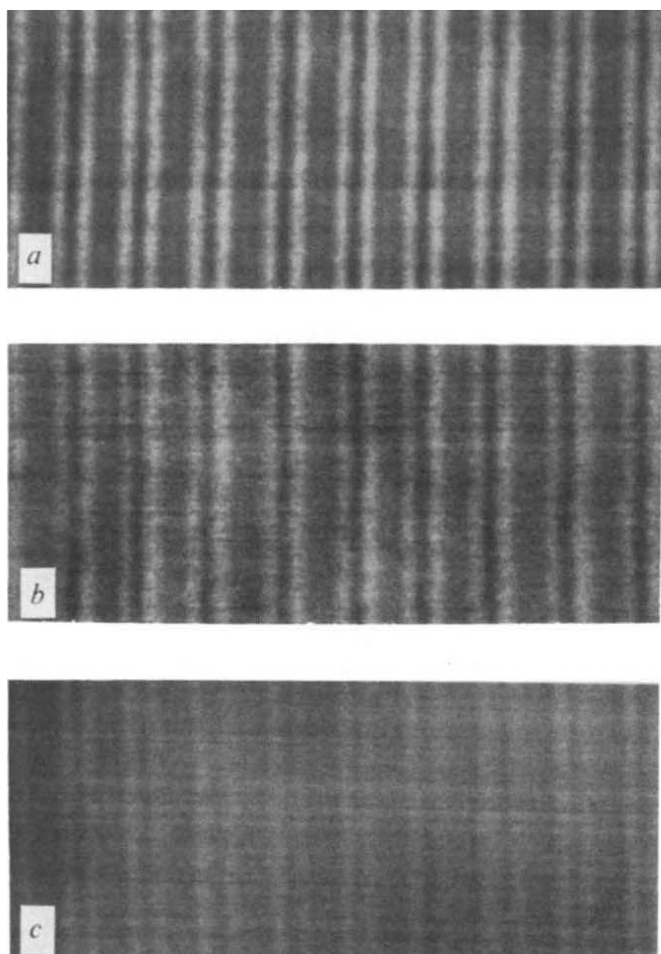


Fig. 3 Z-contrast STEM image of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ at thicknesses of a, 4 nm, b, 27 nm, c, 40 nm.

has been used to image, without resolving the crystal structure, supported-catalyst particles^{2,3} and dopants in solution in silicon^{4,5}. Using a new high-resolution pole piece (spherical-aberration coefficient $C_s \approx 1.3$ mm) and tilting holder in a VG Microscope's HB501 STEM, operating at 100 keV, we have been able to form a probe of <0.25 -nm diameter (measured from images of single uranium atoms supported on a thin carbon film) and have obtained the first images that resolve a crystal lattice while preserving the strong Z-contrast characteristic of high-angle electron scattering.

Figure 1 shows images from cross-sections of $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ and $\text{ErBaCu}_3\text{O}_{7-x}$ viewed parallel to the c (layer) planes; the materials were prepared by mechanical polishing followed by Ar-ion milling. The tripled perovskite cell is clearly visible, the bright lines within the cell corresponding to the planes containing the heavy metal ions. Because of the strong Z-sensitivity, the Y-atom ($Z = 39$) plane shows a lower intensity than the Ba ($Z = 56$) planes (Fig. 1a), whereas in the rare-earth-substituted material the Er ($Z = 68$) plane is brighter than the Ba planes (Fig. 1b). In the limit of a very thin crystal, the Z-contrast image would be given by the projected scattering power of the crystal convoluted with the probe intensity profile. Figure 2 shows calculated intensity profiles across two unit cells for the two superconductors, assuming a gaussian probe of 0.24-nm diameter at $1/e$ intensity and using a screened cross-section expression formulated by Fleischmann⁵, which gives the best fit to experimentally measured cross-sections⁶. This extremely simple calculation reproduces correctly the relative intensities of the various planes seen in Fig. 1. It is analogous to the phase-grating approximation in phase-contrast imaging,

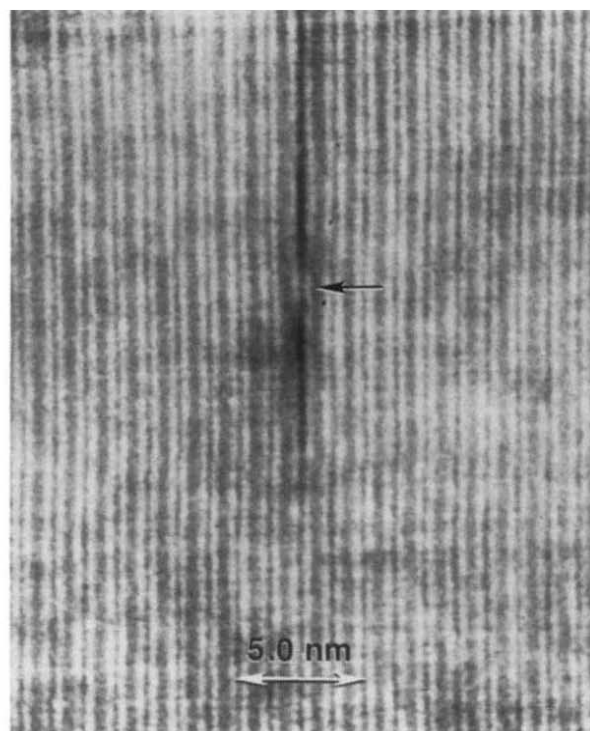


Fig. 4 Z-contrast STEM image of a planar defect in the $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ single crystal. Above the arrow the Ba planes are expanded by $c/6$, but below the arrow they are not.

although it holds to significantly greater thicknesses, because for the first 5 nm or so the electron-channelling phenomenon simply concentrates the electron flux equally onto the various atomic planes in the unit cell. At greater thicknesses the flux oscillates between the various planes, concentrating first on the Ba planes at the expense of the Y plane, and then returning to the Y plane at thicknesses >30 nm. This introduces a thickness dependence in the image, although it is minimized because the image is formed from the integrated scattering throughout the whole sample thickness. Figure 3 shows images at 4, 27 and 40 nm, indicating that the absolute contrast reduces with increasing thickness because of electron absorption, whereas the relative contrast of the Y plane compared to the Ba planes is first reduced and then returns to its original level. This dependence is much less pronounced than in conventional phase-contrast imaging, and the various planes in the cell can be located even in samples 80 nm thick. The dependence of the image on objective-lens defocus is also significantly less than in a conventional phase-contrast image. It serves only to focus the probe on the sample and cannot produce contrast reversals.

In addition to strong chemical sensitivity, the highly localized nature of the electron scattering used in Z-contrast imaging is an advantage when studying defects or interfaces. In a conventional phase-contrast image, the localization is determined by the objective-lens defocus, which also controls contrast and resolution. Figure 4 shows a planar defect in the $\text{YBa}_2\text{Cu}_3\text{O}_{7-x}$ single-crystal cross-section. At this crystal thickness the Y plane is not visible, although the defect is unambiguously seen to lie between the twin Ba planes of the unit cell. The defect expands the Ba planes by $c/6$ (where c is the undistorted lattice parameter) indicating that it is an interstitial-type defect. The dark contrast indicates that the defect must consist of atoms lighter than Y, and the structure immediately suggested is one in which an additional CuO layer is inserted between the Ba planes. This structure has also been proposed from conventional phase-contrast electron microscopy, and is suggested to arise as the response of the material to oxygen sub-stoichiometry⁷ and as a

degradation reaction⁸. Interstitial cation layers and cationic disorders have also been reported^{9,10}, although we see no evidence of these. The $c/6$ expansion is only present in the upper part of Fig. 4 (above the arrow); the lower region of the defect does not dilate the lattice, although dark contrast remains. This lower region cannot correspond to simply an interstitial CuO layer, and we believe that the effect here is due to a degradation reaction, whereas the interstitial CuO plane seen in the upper part of the defect represents the fundamental response of the material to oxygen sub-stoichiometry, as proposed in ref. 7. This interpretation is supported by the recent observation of an ordered array of such defects¹¹.

We are grateful to M. M. J. Treacy for loan of the resolution test sample, to J. O. Ramey for contributions to the crystal growth effort, and to C. W. Boggs and J. T. Luck for sample preparation. This research was sponsored by the Division of Materials Sciences, US Department of Energy, under contract with Martin Marietta Energy Systems, Inc.

Received 18 August; accepted 19 October 1988.

1. Scherzer, O. *J. appl. Phys.* **20**, 20-29 (1949).
2. Treacy, M. M. J. *J. Microsc. Spectrosc. Electron* **7**, 511-523 (1982).
3. Treacy, M. M. J., *J. Microsc.* (in the press).
4. Pennycook, S. J. & Narayan, J. *Appl. Phys. Lett.* **45**, 385-387 (1984).
5. Fleischmann, H. *Z. Naturf.* **15a**, 1090-1096 (1960).
6. Pennycook, S. J., Berger, S. D. & Culbertson, R. J. *J. Microsc.* **144**, 229-249 (1986).
7. Domenges, B., Hervieu, M., Michel, C. & Raveau, B. *Europhys. Lett.* **4**, 211-214 (1987).
8. Zandbergen, H. W., Gronsky, R. & Thomas, G. *Phys. Status Solidi (a)* **105**, 207-218 (1988).
9. Ourmazd, A. *et al. Nature* **327**, 308-310 (1987).
10. Ourmazd, A., Spence, J. C. J., Zuo, J. M. & Li, C. H. *J. Electr. Microsc. Tech.* **8**, 251-262 (1988).
11. Marshall, A. F. *et al. Phys. Rev.* **B 37**, 9353-9358 (1988).

Entropy catastrophe and superheating of crystals

S. Lele*, P. Ramachandra Rao† & K. S. Dubey‡

* Department of Metallurgical Engineering,

† School of Materials Science and Technology, and

‡ Department of Applied Physics, Institute of Technology, Banaras Hindu University, Varanasi-221 005, India

Kauzmann¹ has pointed out that at a temperature (T_k) far below the melting temperature (T_m) of a crystal, the entropies of an undercooled liquid and its corresponding equilibrium crystal tend to become equal. At temperatures below T_k , the crystal would have a greater entropy than its liquid. Such a catastrophe is avoided through a glass transition of the liquid above T_k . Fecht and Johnson² have shown that such a transition should occur for aluminium at $0.23 T_m$. A similar entropy catastrophe can also arise at a temperature, T_i^S , above T_m . Above T_i^S , the crystalline solid is once again expected to have an entropy greater than that of the liquid. Cahn³ has considered the implications of this idea with respect to the superheating of a solid. Here we present an alternative evaluation of the two temperatures of instability, T_k and T_i^S , from experimentally measurable parameters. Results for alkali metals show that the Kauzmann-type entropy catastrophe occurs at about half the absolute melting temperature, whereas we find that the entropy catastrophe as described by Fecht and Johnson² occurs at twice the absolute melting temperature. For most solids, vaporization will probably intervene before the entropy catastrophe temperature above the melting temperature is reached.

Fecht and Johnson² considered the possibility of the isentropic temperature T_i^S coinciding with the temperature at which the volume difference (ΔV) between liquid and solid phases vanishes, for the case of both pure metals and binary alloys. Both ΔV and the entropy difference (ΔS) between these phases are partial derivatives of the Gibbs free energy difference (ΔG), with respect to pressure (P) and temperature (T) respectively. For a pure metal, if ΔS and ΔV simultaneously vanish, ΔG also becomes zero. The temperature at which $\Delta S = \Delta V = 0$ is thus

the point of an equilibrium second-order phase transition from a pure solid to its liquid. In such a case, T_i^S cannot be termed a temperature of instability. However, in the case of a binary alloy, where composition provides a further variable, $\Delta S = \Delta V = \Delta G = 0$ can correspond to a temperature of instability, as discussed by Fecht and Johnson².

Ubbelohde⁴ has discussed previous attempts to evaluate the temperature and pressure at which $\Delta S = \Delta V = 0$. Plots of ΔS and ΔV on melting as a function of pressure appear to extrapolate to a critical pressure at which melting is characterized by $\Delta S = \Delta V = 0$. For potassium this critical pressure and temperature have been estimated as $22,390 \text{ kg cm}^{-2}$ and 498 K (ref. 4). Although pressures up to $30,000 \text{ kg cm}^{-2}$ have been attained experimentally, no phase transition or anomalies have been detected in the pressure/melting-point curves. Better methods for extrapolating ΔS and ΔV are therefore needed for predicting T_i^S and the temperature at which $\Delta V = 0$. In any case, ΔS and ΔV cannot be simultaneously zero during melting at atmospheric pressure because the corresponding isentropic and isenthalpic temperatures do not, in general, coincide. At atmospheric pressure T_i^S can still be a temperature of instability.

We have recently proposed⁵ an expression for the free energy of a liquid with respect to a solid in its stable or metastable states. By a Taylor-series expansion of ΔG about its value at T_m , for constant pressure, we obtain

$$\begin{aligned} \Delta G = & \Delta S^m \Delta T - \Delta C_p^m \left\{ T \ln(T/T_m) + \Delta T \right\} \\ & + T_m \left[\frac{\partial \Delta C_p}{\partial T} \right]_{T_m} \left\{ T \ln(T/T_m) + \Delta T - \frac{1}{2} \Delta T^2 / T_m \right\} \\ & - \frac{T_m^2}{2} \left[\frac{\partial^2 \Delta C_p}{\partial T^2} \right]_{T_m} \left\{ T \ln\left(\frac{T}{T_m}\right) + \Delta T \right. \\ & \left. - \frac{1}{2} \frac{\Delta T^2}{T_m} - \frac{1}{6} \frac{\Delta T^3}{T_m^2} \right\} + \dots \end{aligned} \quad (1)$$

where $\Delta T = T_m - T$, and ΔC_p^m and ΔS^m are the heat capacity and entropy differences respectively between the equilibrium melt and the crystal at the melting temperature. Equation (1) has been tested in cases where heat-capacity data of undercooled phases are available. For example, it has been demonstrated that ΔG of undercooled liquid gallium, magnesium and oxides of boron, silicon and germanium and their equilibrium solid phases can be determined accurately using equation (1) (refs 6 and 7). We have also demonstrated that equation (1) can predict the free energy difference between superheated body-centred cubic and equilibrium face-centred cubic modifications of iron⁶.

Differentiation of equation (1) with respect to T yields an expression for ΔS of the form:

$$\Delta S = \Delta S^m + a \ln T_R - b \Delta T_R + C \Delta T_R^2 - d \Delta T_R^3 \quad (2)$$

where a , b , c and d stand for the constant coefficients involving ΔC_p and its derivatives at T_m , and are given by

$$\begin{aligned} a = & \Delta C_p^m - T_m \left[\frac{\partial \Delta C_p}{\partial T} \right]_{T_m} + \frac{T_m^2}{2} \left[\frac{\partial^2 \Delta C_p}{\partial T^2} \right]_{T_m} - \frac{T_m^3}{6} \left[\frac{\partial^3 \Delta C_p}{\partial T^3} \right]_{T_m} \\ b = & T_m \left[\frac{\partial \Delta C_p}{\partial T} \right]_{T_m} - \frac{T_m^2}{2} \left[\frac{\partial^2 \Delta C_p}{\partial T^2} \right]_{T_m} + \frac{T_m^3}{6} \left[\frac{\partial^3 \Delta C_p}{\partial T^3} \right]_{T_m} \\ c = & \frac{T_m^2}{4} \left[\frac{\partial^2 \Delta C_p}{\partial T^2} \right]_{T_m} - \frac{T_m^3}{12} \left[\frac{\partial^3 \Delta C_p}{\partial T^3} \right]_{T_m} \\ d = & \frac{T_m^3}{18} \left[\frac{\partial^3 \Delta C_p}{\partial T^3} \right]_{T_m} \end{aligned}$$

where $T_R = T/T_m$ and $\Delta T_R = (1 - T_R)$. In equation (2) we have dropped terms involving ΔT_R^4 and higher-order terms. Note that both equations (1) and (2) require only the readily available values of ΔS and ΔC_p at the equilibrium transition temperature and their derivatives at T_m . These derivatives can be obtained

Table 1 Values of T_k and T_i^S for alkali metals derived from equation (3)

Metal	T_m (K)	T_R^i for $\Delta S = 0$	T_k (K)	T_i^S (K)
Li	453.69	0.4953	2.0466	224.7
Na	370.98	0.4089	2.1460	151.7
K	336.0	0.4557	1.9065	153.1
Rb	312.64	0.5353	1.8651	167.4
Cs	301.55	0.5197	2.0272	156.6

using the heat-capacity data of the two phases in their stable regimes.

When $\Delta S = 0$, we obtain

$$\ln T_R^i = -\Delta S^m/a + (b/a)\Delta T_R^i - (c/a)(\Delta T_R^i)^2 + (d/a)(\Delta T_R^i)^3 \quad (3)$$

where T_R^i are the values of the reduced temperature T_R for which $\Delta S = 0$.

If in equation (3) we neglect terms of order $(\Delta T_R^i)^2$ and higher, the straight line representing the right-hand side of the equation intersects the logarithmic curve of the left-hand side at no more than two points. Both of the corresponding values of T_R^i represent temperatures of entropy catastrophe. Inclusion of higher-order terms in the analysis is generally expected to refine the values of T_R^i obtained while still limiting the real and possible solutions to only two. We have adopted this approach for evaluating T_R^i for alkali metals. For these metals all the

required experimental data (ΔS , ΔC_p , $\partial \Delta C_p / \partial T$ and so forth at T_m) are available⁸.

We present our results in Table 1. We have designated the temperature of instability derived from the value of $T_R^i < 1$ as T_k , and that for $T_R^i > 1$ as T_i^S . We find that $T_i^S = (1.99 \pm 0.15) T_m$ and $T_k = (0.48 \pm 0.07) T_m$ for all the metals considered. The values of T_R^i are closely clustered for similar metals. The values of (T_i^S/T_m) are greater, however, than those reported by Fecht and Johnson² for Al, Nb and W. It should be pointed out that our method uses only easily accessible experimental data at the normal melting point. As the actual values of C_p of the solid and liquid phases are used, all contributions to the heat capacity of the solid phases, for example from vacancies, are implicitly accounted for. If T_i^S is twice T_m , in general it is unlikely that T_i^S would ever be attained by superheating a solid at atmospheric pressure. Other instabilities, for example those arising from elastic constants tending to zero, from changes in electronic structure or from vaporization, may intervene. But some of these instabilities can be suppressed by extremely rapid heating with the aid of lasers or electron beams.

Received 25 August; accepted 20 October 1988.

1. Kauzmann, W. *Chem. Rev.* **43**, 219–256 (1948).
2. Fecht, H. J. & Johnson, W. L. *Nature* **334**, 50–51 (1988).
3. Cahn, R. W. *Nature* **334**, 17–18 (1988).
4. Ubbelohde, A. R. *The Molten State of Matter* (Wiley, New York, 1978).
5. Lele, S., Dubey, K. S. & Ramachandra Rao, P. *Curr. Sci.* **54**, 994–995 (1985).
6. Ramachandra Rao, P., Dubey, K. S. & Lele, S. *Acta Metall.* (in the press).
7. Dubey, K. S., Ramachandra Rao, P. & Lele, S. *Phys. Chem. Glasses* (submitted).
8. Barin, I., Knacke, O. & Kubaschewski, O. *The Thermochemical Properties of Inorganic Substances*, and Supplement (Springer, Berlin, 1973 & 1977).

Long-range transport of giant mineral aerosol particles

P. R. Betzer*, K. L. Carder*, R. A. Duce†, J. T. Merrill†, N. W. Tindale†, M. Uematsu†§, D. K. Costello*, R. W. Young*, R. A. Feely‡, J. A. Breland*, R. E. Bernstein* & A. M. Greco*

* Department of Marine Science, University of South Florida, St Petersburg, Florida 33701, USA

† Center for Atmospheric Chemistry Studies, Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882, USA

‡ National Oceanic and Atmospheric Administration, Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, Washington 98115, USA

Several recent studies have shown that large quantities of mineral dust from eastern Asia are transported through the atmosphere to the North Pacific each spring^{1–5}. The paucity of information on mineral fluxes during individual dust events prompted a coordinated effort, Asian Dust Input to the Oceanic System (ADIOS), which simultaneously measured mineral fluxes in the atmosphere and upper water column during such an event. In March 1986 a major dust outbreak in China moved over the North Pacific Ocean and was detected downstream using changes in particle number, size and composition. Most striking was the presence of 'giant' (>75-μm) silica minerals found in atmospheric as well as water-column samples at the ADIOS sampling site (26° N, 155° W). Their appearance more than 10,000 km from their source cannot be explained using currently acknowledged atmospheric transport mechanisms. Furthermore, the large wind-blown minerals that dominated our samples are extremely rare in the long-term sedimentary record in the North Pacific.

Concentrations and fluxes of mineral particles were deter-

mined in the atmosphere and upper water column for a 27-day period in March/April 1986 at a station in the North Pacific (26° N, 155° W) aboard the R/V *Moana Wave*. Atmospheric sampling was also undertaken from platforms on the windward coasts of Oahu and Midway (Fig. 1). Procedures for collection and analysis of atmospheric aerosol and total-deposition samples for total aluminium, an indicator of aluminosilicate minerals, have been described previously^{4,5}. Mineral fluxes in the water column were estimated with free-drifting sediment traps⁶, some of which were fitted with programmable sample changers which isolated a sample every four hours⁷. Traps were deployed at depths of 37, 127, 400, 900, 2,200 and 3,500 metres.

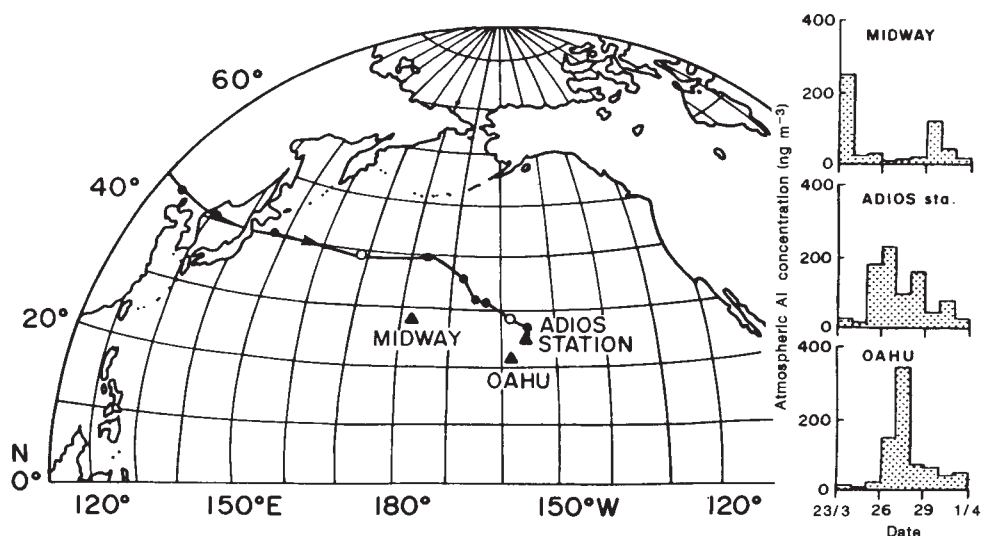
Atmospheric and water-column samples were analysed using computer-controlled scanning electron microscopy (SEM) and energy-dispersive X-ray analysis (EDXA)⁸. In brief, each of the particles was sized (length/width) and then analysed for 5 s. Based on the proportion of key elements such as Si, Al, Fe, Ti, Mg and Ca, the particles were sorted into chemical macroclasses as a function of size.

According to routine surface meteorological reports from China, a major dust outbreak occurred in 1986 between 15 and 17 March. Relatively low atmospheric aluminium concentrations were observed initially at the ADIOS site on 23 March, but were followed by substantial increases. At Midway our sampling was initiated during and/or after the arrival of the dust from Asia (Fig. 1). The dust pulse reached the ADIOS Station on 25 March, at which time atmospheric aluminium concentrations increased tenfold. A similar concentration change was observed at Oahu one day later. Isentropic back trajectories at 300 K (ref. 9) for air reaching the ADIOS Station on 25 March are consistent with the movement of material from Asia to the ADIOS Station in the open Pacific Ocean with a transit time of 9–11 days.

Atmospheric total-deposition samples collected at the ADIOS site during this dust event showed considerable increases in the particle number flux and mass flux for particles of diameter >75 μm compared with the pre-dust input values (Table 1). Associated with the increased flux was a marked change in the particle size distribution of the material deposited. Before the

§ Present address: Department of Marine Sciences and Technology, Hokkaido Tokai University, Minami-sawa, Minami-ku, Sapporo, 005 Japan.

Fig. 1 Map of the sampling region and dust trajectory. Triangles represent sampling stations in the North Pacific Ocean. Atmospheric aluminium concentrations observed during the late March 1986 Asian dust event are shown on the right. An isentropic trajectory at 300 K, reaching the ADIOS station on 26 March 1986, is also shown. The air mass crossed the central-northern North Pacific, moving towards the east at ~4 km altitude, and then descended as it moved into the marine boundary layer. Dots along this trajectory represent 24-h travel times, with open circles at 5-day intervals.



arrival of the dust pulse, materials deposited at sea were dominated both numerically (>99%) and by mass (>97%) by particles of <75 μm in area-equivalent diameter (equal to $2\sqrt{\text{cross-sectional area}/\pi}$). Although the dust pulse that appeared on 25 March at our open-ocean station was characterized by a small increase in the total mass flux, there was a distinctive change in the particle size distribution resulting from the presence of 'giant' particles. These were primarily quartz grains, several of which had area-equivalent diameters >200 μm . The enhanced total mass flux and input of giant particles continued through the next two atmospheric sampling periods (Table 1).

A corresponding increase in mineral fluxes was detected in the upper water column by the drifting sediment traps. Two sequential 4-h samples collected from a single trap at the 37-m horizon between 25 and 26 March showed distinct changes in the abundance, size and chemical character of particulate material (Table 2). Initially, the particulate material isolated in

the trap consisted of relatively small, calcium-dominated particles. The latter sample also had a large number of calcium-dominated particles, but, in sharp contrast to the earlier sample, also contained large silica-bearing particles and increased amounts of various clays and iron-rich materials (Table 2). Four representative giant quartz grains isolated from this trap are shown in Fig. 2. The large silica, clay and iron-rich particles noted in the water column were similar both in size and in composition to the atmospheric input material. The flux of non-biogenic material during the latter collection period ($447 \mu\text{g m}^{-2} \text{d}^{-1}$) is significantly greater than that noted in the earlier period ($43 \mu\text{g m}^{-2} \text{d}^{-1}$) and is also in good agreement with the range of values from the coincident two-day atmospheric-deposition samples ($300\text{--}560 \mu\text{g m}^{-2} \text{d}^{-1}$).

We stress that the detection of large mineral particles at the ADIOS Station did not result from sample contamination during collection or processing. The giant, chemically distinct particles appeared simultaneously in atmospheric and shallow marine

Table 1 Atmospheric deposition of particles to the ocean surface

Sample	Area-equivalent diameter range (μm)					Total
	20-50	50-75	75-100	100-200	>200	
08:00, 23/3-08:10, 24/3						
Total particle no. flux	9,776	716	44	0	0	10,536
Quartz particle no. flux	186	22	0	0	0	208
Total mass flux	135	112	8	0	0	255
Quartz mass flux	1	4	0	0	0	5
08:10, 24/3-08:30, 26/3						
Total particle no. flux	1175	171	123	68	6	1543
Quartz particle no. flux	243	132	97	68	6	546
Total mass flux	16	31	76	159	16	298
Quartz mass flux	5	26	63	159	16	269
08:30, 26/3-08:30, 28/3						
Total particle no. flux	3,645	458	156	119	8	4,386
Quartz particle no. flux	1,138	350	156	111	8	1,763
Total mass flux	55	76	96	225	109	561
Quartz mass flux	21	66	96	218	109	510
08:30, 28/3-08:30, 30/3						
Total particle no. flux	5,658	979	518	305	38	7,498
Quartz particle no. flux	2,462	808	499	286	38	4,093
Total mass flux	91	185	285	791	482	1,834
Quartz mass flux	53	161	283	775	482	1,754

Particle number flux in particles $\text{m}^{-2} \text{d}^{-1}$; mass flux in $\mu\text{g m}^{-2} \text{d}^{-1}$. The figures consider only particles >15 μm in area-equivalent diameter, and do not include sea-salt aerosol particles, specifically particles enriched in NaCl. The number of single particles analysed in each sample (not including sea-salt particles): 23/3-24/3 1,729; 24/3-26/3 897; 26/3-28/3 2,050; 28/3-30/3 2,046.

Table 2 Transition in particle fluxes in the upper water column of the North Pacific Ocean

Response to input of atmospheric mineral debris Sampling period: 16:00 to 20:09 on 25 March Total non-biogenic mass flux = 43 $\mu\text{g m}^{-2} \text{d}^{-1}$							
Chemical category	Particle fluxes* by area-equivalent diameter ranges (μm)					Total number	Mass flux†
	10–25	25–50	50–100	100–200	>200		
Ca-rich	863	592	197	73	0	1,725	128
Ca only	246	98	271	122	24	791	410
Ca/Si	123	24	0	0	0	147	1
Al/Si/Ca	24	0	24	0	0	48	3
Al/Si/Fe	0	0	0	0	0	0	0
Al/Si/Mg/Fe	127	148	24	0	0	344	8
Al/Si/K/Fe	24	73	0	0	0	97	3
Al/Si/clay	73	24	0	0	0	97	0
Ti-rich	0	0	0	0	0	0	0
Fe-rich	196	49	0	0	0	245	7
Fe only	295	49	24	0	0	368	12
Si-rich	123	49	0	0	0	172	3
Si only	345	74	24	0	0	443	6
Particle number flux	2,439	1,180	564	195	24	4,447	581
Non-biogenic mass flux	7	18	18	0	0		43

Sampling period: 20:09 to 00:18 on 25/26 March Total non-biogenic mass flux = 447 $\mu\text{g m}^{-2} \text{d}^{-1}$							
Chemical category	Particle fluxes* by area-equivalent diameter ranges (μm)					Total number	Mass flux†
	10–25	25–50	50–100	100–200	>200		
Ca-rich	806	339	84	0	0	1,229	29
Ca only	1,103	764	169	0	0	2,036	42
Ca/Si	508	254	84	0	0	846	15
Al/Si/Ca	42	0	42	0	0	84	1
Al/Si/Fe	1,187	254	42	0	0	1,483	18
Al/Si/Mg/Fe	3,013	2,122	763	127	0	6,025	205
Al/Si/K/Fe	1,695	465	42	0	0	2,202	16
Al/Si/clay	1,526	467	0	0	0	1,993	11
Ti-rich	84	84	0	0	0	168	3
Fe-rich	508	382	42	0	0	932	23
Fe only	593	339	126	0	0	1,058	67
Si-rich	933	297	42	42	0	1,314	57
Si only	1,188	467	42	0	0	1,697	31
Particle number flux	13,186	6,234	1,478	169	0	21,067	518
Non-biogenic mass flux	45	112	217	73	0		447

Number of particles analysed in each sample: 16:00–20:09 on 25/3: 719; 20:09–00:18 on 25–26/3: 1,917. Local times are used. Masses calculated as the sum of individual particle masses derived from size and chemical composition data obtained from SEM-EDXA analyses. The two-day atmospheric deposition samples indicate a mineral aerosol flux of 300–560 $\mu\text{g m}^{-2} \text{d}^{-1}$.

* Particle number flux in particles $\text{m}^{-2} \text{d}^{-1}$.

† Mass flux in $\mu\text{g m}^{-2} \text{d}^{-1}$. The first two chemical categories represent the biogenic contribution.

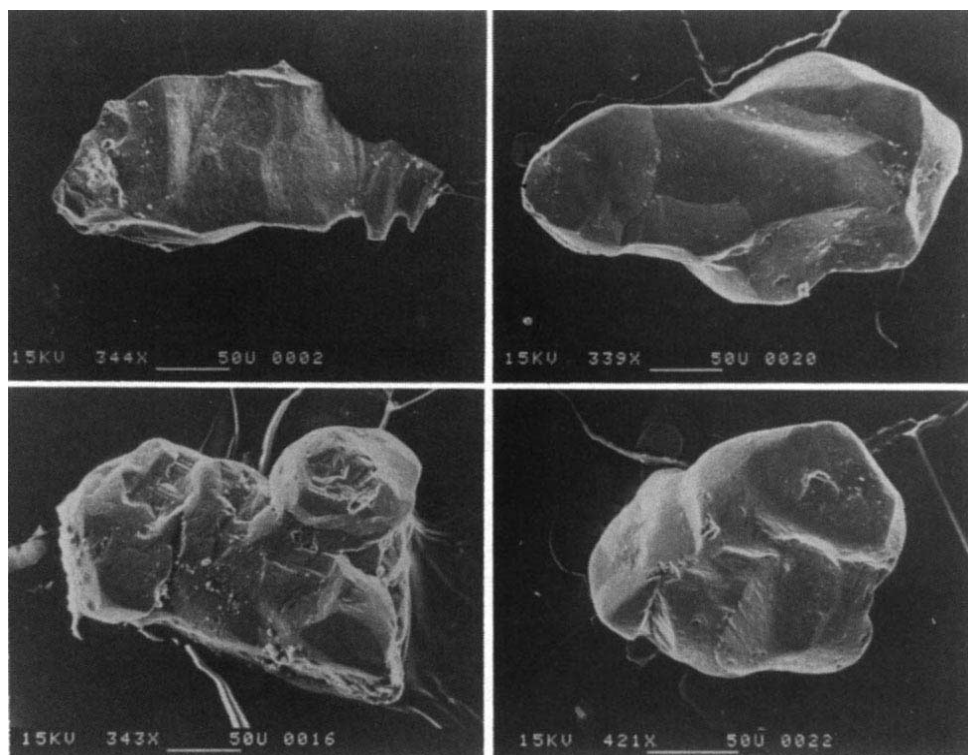
samples. Furthermore, in the water column they were detected in two different sediment traps, deployed several kilometres apart, which were sampling the 37-m horizon on 25 and 26 March, and also in other separate sediment traps sampling at 900, 2,200 and 3,500 m; they were also isolated from 30-litre Niskin bottles and collected from *in situ* filtrations of suspended matter¹⁰. In addition, the calculated settling times of the giant silica-bearing materials trapped at 900, 2,200 and 3,500 m also indicate that the arrival of the dust event at our open-ocean location lies somewhere between 20:00, 25 March and 00:18, 26 March. It is highly unlikely that such a consistent pattern would result from artefacts or random contamination. Finally, the composition and size ranges of these giant particles are consistent with their proposed source, specifically China's extensive deserts and/or loess deposits^{11,12}.

Nonetheless, it is extremely difficult to account for the appearance of giant particles over 10,000 km from their source region. The atmospheric-deposition velocity estimates of Windom¹³ allow 20- μm diameter quartz particles to travel ~7,000 km if their initial injection height is 8 km. Gillette¹⁴ invokes a re-suspension probability in which an exponentially decreasing

fraction of particles in each settling velocity interval can be carried further and further; however, the parameters he uses are for the atmospheric boundary layer and are not germane to long-range transport. Even with the use of sophisticated meteorological models, including full aerosol microphysics, it is not possible to account for even 30- μm diameter particles from the Sahara travelling a distance of 1,000 km (ref. 15). Although some workers have invoked agglomeration near the sampling location to explain the presence of large particles^{15,16}, Prospero *et al.*¹⁷ and Carder *et al.*¹⁸ detected individual large (>20- μm diameter) particles that were carried more than 4,000 km from their Saharan source. The individual large particles reported here are considerably larger than those detected in refs 17 and 18, and require a re-suspension mechanism. It is our hypothesis that isolated convective storms along the particles' path provide the required uplift.

Although we are unable to specify the physical mechanisms involved in their transport, the data show the considerable importance of the giant particles to the total mass flux in this event (Tables 1 and 2). In our atmospheric samples, over 83% of the mass flux for particles larger than 20 μm was due to giant

Fig. 2 Scanning electron micrographs of four giant quartz particles collected from a sediment trap deployed at a depth of 37 m. The white bar in each photograph represents a distance of 50 μm . Energy-dispersive X-ray analysis of these particles is consistent with their being classified as quartz. These particles have length:width ratios ranging from nearly spherical to 2.5:1, and their topographies suggests that these grains have substantial thickness. In combination, their size and shape are indicative of rapid settling velocities.



(>75- μm diameter) particles during 24–30 March (Table 1). Quartz particles alone comprise 95% of the mass influx in the >75- μm particle range during the dust event (24–26 March). This can be contrasted with the pre-dust data (23–24 March), where less than 4% of the mass input is due to giant particles. During this same period, quartz particles of diameter >20 μm account for only ~2% of the mass input.

When considered in the context of data from the SEAREX Asian Dust Network and routine meteorological observations, the event discussed here appears unremarkable. For example, the peak atmospheric concentrations of dust observed for this event were well within the usual range for dust events observed at North Pacific island sites⁴. The transport meteorology follows the pattern usually observed⁹, and the number of routine meteorological reports of dust during the source event, although higher than the seasonal average peak value by ~20%, are not exceptional. Furthermore, the estimated atmospheric deposition for the entire event (3,000 $\mu\text{g m}^{-2}$) is less than 1% of the annual average aerosol flux measured in this area⁵. What is clearly remarkable about this event is the presence of such large mineral debris which was carried over 10,000 km from the Asian continent to our open-ocean sampling site. Evidence from other events encountered during the 1986 ADIOS expedition indicates that this event is not unique. That is, although the flux of particles with diameters >75 μm is greater in this event than in others, particles with diameters >50 μm are common in all these events. Large-diameter quartz particles of aeolian origin are found in bottom sediments over large portions of the Central North Pacific Ocean¹⁹, but their proportions are far smaller than those noted in this work. Furthermore, in contrast to our data for both atmosphere and water column, surface sediments from two different sites near the ADIOS Station show no quartz particles of >55- μm diameter¹⁹. Thus, when viewed in the context of the long-term sedimentary record (~2,000 yr) at the bottom of an ocean whose non-biogenic sedimentary materials are essentially all of aeolian origin²⁰, this event is certainly not typical.

It is possible that the actions of man have altered wind erodibility and have thereby modified the size distribution of the mineral aerosol particles at the source; anthropogenic alter-

ations in dust characteristics have been noted elsewhere²¹. Additional research is needed to understand the mechanisms responsible for the long-range transfer of giant mineral aerosols.

The officers and crew of the R/V *Moana Wave* and the personnel of the University of Hawaii Marine Research Center did an excellent job of facilitating this work. We especially note the superb efforts of Captain W. Leonard, Deck Technician Curtis See, and Van Chisholm. We appreciate the data of Paul Dauphin, the comments of Margaret Leinen, and help from the National Center for Atmospheric Research Computer Facility. This work was supported by grants from the NSF to the University of South Florida, the University of Rhode Island and the University of Hawaii. The National Oceanic and Atmospheric Administration, through their PMEL facility, also supported this research.

Received 10 August; accepted 31 October 1988.

1. Duce, R. A., Unni, C. K., Ray, B. J., Prospero, J. M. & Merrill, J. T. *Science* **209**, 1522–1524 (1980).
2. Darzi, M. & Winchester, J. W. *J. geophys. Res.* **87**, 1251–1258 (1982).
3. Parrington, J. R., Zoller, W. H. & Aras, N. K. *Science* **220**, 195–197 (1983).
4. Uematsu, M. *et al. J. geophys. Res.* **88**, 5343–5352 (1983).
5. Uematsu, M., Duce, R. A. & Prospero, J. M. *J. Atmos. Chem.* **3**, 123–138 (1985).
6. Betzer, P. R. *et al. Deep Sea Res.* **31**, 1–11 (1984).
7. Costello, D. K., Young, R. W., Carder, K. L. & Betzer, P. R. *Eos* **67**, 899 (1986).
8. Johnson, D. L. in *Proc. Conf. Toxic Hazardous Chemicals* (ed. Pilon, K.) 161–169 (Central New York Water Quality Program, New York, 1980).
9. Merrill, J. T., Bleck, R. & Avila, L. *J. geophys. Res.* **90**, 12927–12936 (1985).
10. Schatzberg, P., Adema, C. M., Thomas, W. M. & Mangum, S. R. in *Oceans '86 Conference Record: Science-Engineering-Adventure* Organotin Symposium, Mar. Tech. Soc., Washington, DC, Vol. 4, 1155–1159 (IEEE Publ. Ser., New York, 1986).
11. Sheng, L. T., Fei, G. X., Sheng, A. Z. & Xiang, F. Y. in *Desert Dust: Origin, Characteristics, and Effect on Man* (ed. Péwé, T. L.) Spec. Paper 186, 149–157 (Geol. Soc. Am., Boulder, 1981).
12. *Loess in China* (eds Wang, Y.-y. & Zhang, Z.-h.) (Shaanxi People's Art Publishing House, Xian, China, 1980).
13. Windom, H. L. *Bull. geol. Soc. Am.* **80**, 761–782 (1969).
14. Gillette, D. A. in *Desert Dust: Origin, Characteristics, and Effects on Man* (ed. Péwé, T. L.) Spec. Paper 187, 11–26 (Geol. Soc. Am., Boulder, 1981).
15. Westphal, D. L., Toon, O. B. & Carlson, T. N. *J. geophys. Res.* **92**, 3027–3049 (1987).
16. Schutz, L. in *Saharan Dust* (ed. Morales, C.) 267–277 (Wiley, New York, 1979).
17. Prospero, J. M., Bonatti, E., Schubert, C. & Carlson, T. N. *Earth planet. Sci. Lett.* **9**, 287–293 (1970).
18. Carder, K. L., Steward, R. G. & Betzer, P. R. *J. geophys. Res.* **91**, 1055–1066 (1986).
19. Dauphin, J.-P. thesis, (Univ. Rhode Island, 1983).
20. Blank, M., Leinen, M. & Prospero, J. M. *Nature* **314**, 84–86 (1985).
21. Middleton, N. J. *Nature* **316**, 431–434 (1985).

Discovery of a second Ordovician meteorite using chromite as a tracer

Jan Olov Nyström*, Maurits Lindström†
& Frans E. Wickman†

* Swedish Museum of Natural History, S-10405 Stockholm, Sweden

† Department of Geology, University of Stockholm,
S-10691 Stockholm, Sweden

The small number of known fossil meteorites¹ owe their discovery to the preservation of a characteristic composition or structure. These features are easily changed beyond recognition by diagenesis and other processes at low temperature after a meteorite becomes incorporated into sediments, because the meteoritic minerals are generally susceptible to even weak alteration. The mineral chromite, however, is an exception. Here we report the finding of a strongly altered fossil stony meteorite identified by the composition of its relic chromite, and suggest that chromite chemistry can be used as a basis for a systematic search for fossil meteorites.

The discovery of Brunflo, a structurally well-preserved chondrite in Ordovician limestone^{1,2}, demonstrated that clasts in limestone of hemipelagic type should be examined as potential meteorites. A remnant of one such clast was found by one of us (M.L.) in a slab dumped at the Österplana quarry (lat. 58°35' N, long. 13°26' E) at Kinnekulle, southern Sweden. It is shown below to be a fossil stony meteorite, referred to in this letter as Österplana.

The clast (Fig. 1) occurs in reddish-brown middle Ordovician limestone which contains a rich conodont fauna indicating an age at the Arenigian/Llanvirnian boundary (~480 Myr before present). It is angular with a pentagonal outline, ~4 × 5 cm in cross-section and composed of a rather friable greyish-green material that largely had fallen out before the discovery. A 2–6 cm wide aureole of light greenish-grey colour surrounds it. Megascopically the clast lacks a distinctive structure, but under the microscope a few diffuse rounded bodies with diameters ~1 mm which might represent former chondrules can be discerned. Irregular streaks of semi-opaque unidentified material constitute a discontinuous rim towards the limestone.

The predominating minerals in what is left of the clast are calcite, a green Cr-bearing layer silicate and barite. Apatite forms a 0.7 × 0.4 mm large body, and tiny grains of chromite and a TiO₂ mineral are widespread in small amounts. This composition has little resemblance to the mineralogy of any meteorite besides Brunflo, with which on the other hand there is a striking similarity except for the lack of Co-Ni arsenides and Cr sulphides.

Most of the chromite grains in the clast are anhedral; a tendency towards euhedral form can be observed locally (Fig.



Fig. 1 The fossil stony meteorite Österplana and the surrounding light-coloured aureole in Ordovician limestone; the aureole is narrower across the bedding plane.

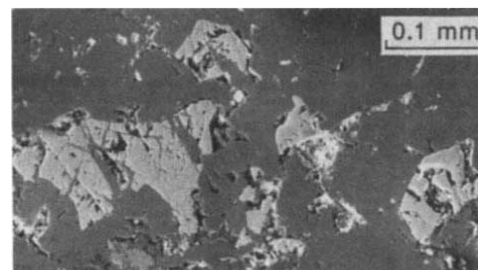


Fig. 2 Scanning electron microscopy photograph of a chromite grain cluster (light) embedded in barite in Österplana (the white disturbance in the centre is a charge effect). The grain at right is partly euhedral.

2). Their typical size range is 0.05–0.1 mm, and they occur singly or in clusters which might be the remnants of larger grains. The reflectivity is uniform, and there is no sign of zoning, although analysis of some grains indicates a local depletion in iron. Such grains are excluded from the average composition given in Table 1.

Chemically, the chromite in Österplana is very similar to that in Brunflo and in some chondrites (Table 1). In fact, with respect to chromite, ordinary chondrites³ as a group (including some achondrites) are more similar to Österplana than any other category of rocks (Figs 3 and 4) and we have found no example of terrestrial chromite in the literature compositionally like that in the clast. Chromites in mafic and ultramafic rocks^{4–13} are significantly different, being poorer in Ti and as a rule richer in Al and Mg; the Mn content also tends to be lower. A depletion in Al and Mg is typical of altered chromites and may also be displayed by metamorphosed ones, sometimes with values in the same range as Österplana^{14,15}. However, Ti is unmodified or only slightly changed during the alteration and can therefore be used to discriminate between altered chromite in mafic-ultramafic rocks and chromite in chondrites. Chromites with Ti, Cr and Al contents approaching the Österplana values are reported in some kimberlites and ultramafic nodules in volcanic rocks^{16,17}, but the extraterrestrial chromites differ by having lower Mg:Fe ratios. The composition of the probably corroded though otherwise unaltered chromite relics strongly suggests that Österplana is of extraterrestrial origin. A moderate corrosion

Table 1 Electron microprobe analyses of chromite in Österplana and Brunflo, and similar chromite in a chondrite fall

	Österplana	Brunflo	Pultusk
Cr ₂ O ₃	56.6 ± 0.47	57.1	55.9
Al ₂ O ₃	5.6 ± 0.54	6.5	6.0
V ₂ O ₃	0.73 ± 0.07	0.74	0.65
TiO ₂	3.3 ± 0.15	1.75	2.72
FeO	29.1 ± 0.40	30.9	30.1
MgO	3.3 ± 0.33	1.84	2.72
MnO	0.71 ± 0.09	0.45	1.03
ZnO	0.36 ± 0.23	0.56	1.09
Total	99.70	99.84	100.21

Mean values (and standard deviation for Österplana) based on 14 analyses of eight grains for Österplana, 53 analyses for Brunflo (H4–H5)¹, and four analysed grains for Pultusk (H5)³.

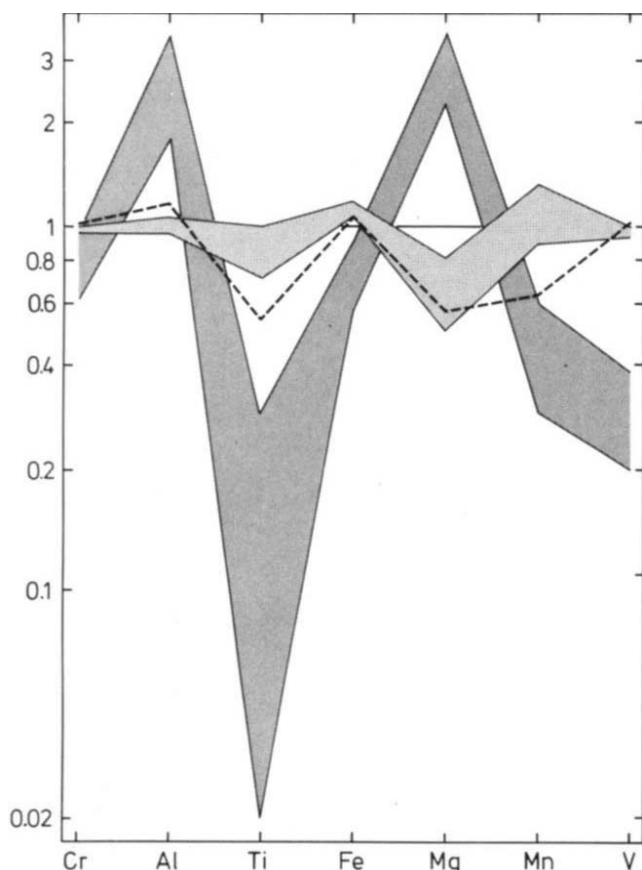


Fig. 3 Geochemical patterns for chromite from Brunflo¹ (broken line), ordinary chondrites (light stippled band, made up of averages for H, L and LL chondrites³), and mafic-ultramafic terrestrial rocks (dark stippled band, composed of averages for ten complexes⁴⁻¹³, the V data are based on only two complexes), normalized against the average composition of the Österplana chromite.

of the 'coarse chromite' grains normally found in stony meteorites¹⁸ would produce grain clusters with dimensions similar to those in Österplana.

Clasts are unexpected in the Ordovician limestones at Kinnekulle which were deposited in a very tranquil environment¹⁹. Furthermore, a transported clast of angular shape is unlikely to occur alone, and the small outcrops of older ultramafic rocks known in the region contain accessory chromite of 'normal' composition. The presence of diffuse rounded bodies in the clast resembling poorly preserved chondrule structures in Brunflo supports the chemical indication that Österplana is an altered chondrite. The similarity in mineralogy and setting between Österplana and Brunflo is consistent with this interpretation. The method of using relic chromite grains of unusual composition in an anomalous setting (an early Precambrian banded iron formation) to infer an extraterrestrial origin has been attempted before²⁰. However, these grains "do not have a composition typical for meteoritic chromites".

It is unlikely that Österplana is an LL chondrite because its chromite is too low in Fe and somewhat high in Mg (see Table 1 in this paper and Tables 2-4 in ref. 3). The question of whether it belongs to the H, L or some other as yet undefined group cannot be answered at present. Österplana fell more than 500 km south of Brunflo and was found in ~5 Myr older strata; thus the two meteorites cannot be related to each other.

The present mineralogy of Österplana is the result of metasomatic reactions that took place after its fall into calcareous mud. Palaeothermal analysis of conodonts from the limestone hosting Österplana give a colour alteration index of 1-1.5, corresponding

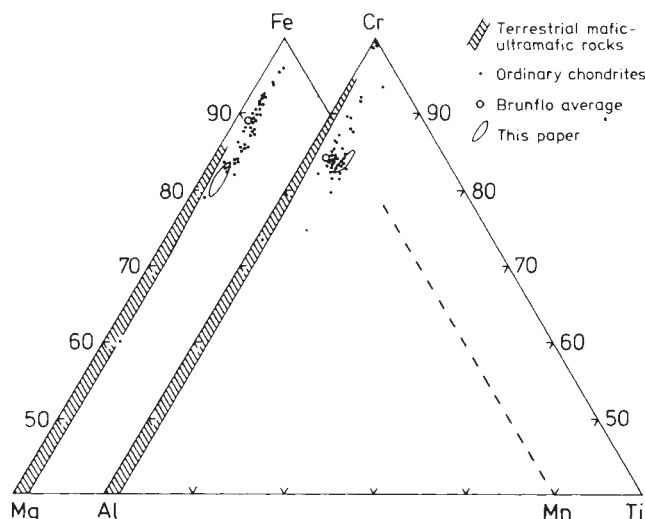


Fig. 4 Compositional variation of chromite in mafic-ultramafic terrestrial rocks⁴⁻¹³ and ordinary chondrites³ (anomalous compositions represent chondrites of subgroups 3). Chromites from ores²⁴ plot in the mafic-ultramafic field, and those from achondrites²⁵ resemble chondrite chromites, though with a trend towards higher Al values.

to a temperature less than 100 °C²¹. The texturally much better preserved Brunflo meteorite, which is surrounded by a more complex multizone aureole, experienced a higher temperature (200-300 °C¹). An explanation for this anomaly is the fact that Österplana fell in an area with slower sedimentation, so the early phase of diagenesis was longer than in the Brunflo area.

The search for meteorites in Antarctica has led to a large number of finds. Clast-free sediments accumulated at a slow rate constitute another environment where meteorites—even if altered—would be conspicuous. The apparent lack of fossil meteorites in the large volumes of coal and limestone that have been mined and quarried²² suggests that clasts of possible meteoritic origin in coal mines and quarries must have been overlooked or ignored, rather than being non-existent. This is demonstrated by the circumstances leading to the discovery of Brunflo and Österplana: both were sent to dumps; a part of the former—by chance situated in the centre of a slab—was saved as a curiosity¹, and the latter would never have been studied without the example of Brunflo. We believe that a systematic search in appropriate sedimentary rocks will increase the number of known fossil meteorites. A search should also be made for tiny chromites in the boundary clays considered to be related to major impact events. Relic unmelted grains of chromite occur in ablation spheres derived from stony meteoroids that have been extracted from deep-sea sediments²³. If any material survives from the meteorites responsible for major impacts refractory chromite is the most likely mineral, and its chemistry might be diagnostic of the nature of the impacting body.

Received 28 September; accepted 30 October 1988.

1. Thorslund, P., Wickman, F. E. & Nyström, J. O. *Lithos* **17**, 87-100 (1984).
2. Thorslund, P. & Wickman, F. E. *Nature* **289**, 285-286 (1981).
3. Bunch, T. E., Keil, K. & Snetsinger, K. G. *Geochim. cosmochim. Acta* **31**, 1569-1582 (1967).
4. Bevan, J. C. & Rodgers, K. A. *Mineralog. Mag.* **41**, 391-394 (1977).
5. Bichan, R. *Econ. Geol. Monograph* **4**, 95-113 (1969).
6. Cameron, E. N. *J. Petrol.* **19**, 437-462 (1978).
7. Challis, G. A. *J. Petrol.* **6**, 322-364 (1965).
8. Himmelberg, G. R. & Loney, R. A. *Bull. geol. Soc. Am.* **84**, 1585-1600 (1973).
9. Loney, R. A., Himmelberg, G. R. & Coleman, R. G. *J. Petrol.* **12**, 245-309 (1971).
10. MacGregor, I. D. & Smith, C. H. *Can. Mineral.* **7**, 403-412 (1963).
11. Malpas, J. & Strong, D. F. *Geochim. cosmochim. Acta* **39**, 1045-1060 (1975).
12. Sinton, J. M. *J. Petrol.* **18**, 216-246 (1977).
13. Summons, T. G., Green, D. C. & Everard, J. L. *Econ. Geol.* **76**, 505-518 (1981).
14. Ahmed, Z. & Hall, A. *Chemie Erde* **40**, 309-339 (1981).

15. Evans, B. W. & Frost, B. R. *Geochim. cosmochim. Acta* **39**, 959-972 (1975).
16. Tompkins, L. A. & Haggerty, S. E. *Contr. Miner. Petrol* **91**, 245-263 (1985).
17. Smith, J. V. & Dawson, J. B. *Phys. Chem. Earth* **9**, 309-322 (1975).
18. Ramdohr, P. *The Opaque Minerals in Stony Meteorites* (Elsevier, Amsterdam, 1973).
19. Jaanusson, V. *Palaeont. Contr. Univ. Oslo* **279**, 1-10 (1982).
20. Appel, P. W. U. *J. Geol.* **87**, 573-578 (1979).
21. Epstein, A. G., Epstein, J. B. & Harris, L. D. *US geol. Surv. prof. Pap.* **995**, 1-27 (1977).
22. Mason, B. *Meteorites* (Wiley, New York, 1962).
23. Blanchard, M. B., Brownlee, D. E., Bunch, T. E., Hodge, P. W. & Kyte, F. T. *Earth planet. Sci. Lett.* **46**, 178-190 (1980).
24. Haggerty, S. E. *Oxide Miner., Rev. Mineral* **3**, Hg101-Hg300 (1976).
25. Bunch, T. E. & Keil, K. *Am. Mineral.* **56**, 146-157 (1971).

Herbivory as a selective agent on the timing of leaf production in a tropical understory community

T. Mitchell Aide

Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA

In highly seasonal environments (for example, temperate deciduous forest or tropical dry forest) plant phenologies are often constrained by physiological stress caused by seasonal variation in temperature and moisture¹. Where climatic fluctuations are less pronounced (for example, tropical marine or lowland wet forest communities) determinants of phenology may include biotic selective pressures such as herbivores, predators, competitors, pollinators or dispersers. Temporal patterns of leaf production vary considerably between tropical forest species, but few species produce leaves continuously throughout the year². Even in aseasonal tropical forests, most species show synchronous peaks in leaf production². Results from this one-year study suggest that herbivores may influence leaf phenology of shade-tolerant understory plants in a Panamanian lowland moist forest. Leaves produced during the dry season or in synchronous flushes receive significantly less damage from herbivores than leaves produced during the wet season or out of synchrony with conspecifics.

For a wide range of tropical and temperate plant species, the average biomass lost to herbivores is greater than the biomass allocated to reproduction³. In response to the strong selective pressures of herbivores, plants have evolved a variety of chemical and structural defences^{4,5}. In many cases these defences are most effective against herbivores of mature leaves, resulting in the majority of damage occurring during the period of leaf expansion⁶. Although expanding young leaves are often chemically defended, the high levels of water and nitrogen and low toughness make them especially vulnerable to herbivores⁶⁻¹⁰. Because leaves expand for a period of only weeks, the timing of leaf production may influence levels of damage.

It has been suggested that two phenological mechanisms may reduce damage to expanding leaves: (1) producing leaves when herbivore abundance is low, or (2) satiating herbivores with highly synchronous flushes^{6,8,11-13}. Trees experimentally defoliated in a Costa Rican dry forest responded by producing a second cohort of leaves¹⁴. The cohort produced 7-10 days later in the wet season and out of synchrony with the population, received significantly higher levels of damage. This suggests the importance of synchronous leaf production as a possible mechanism for escaping or satiating young-leaf herbivores. Rockwood¹⁴ also observed trees 'anticipating' the wet season and producing leaves at the end of the dry season before the appearance of the herbivores, thereby seasonally escaping damage. Although the majority of tropical species produce leaves synchronously, pioneer or gap species growing in high light often produce leaves continuously. McKey¹³ has predicted that the young leaves of these species are unlikely to escape

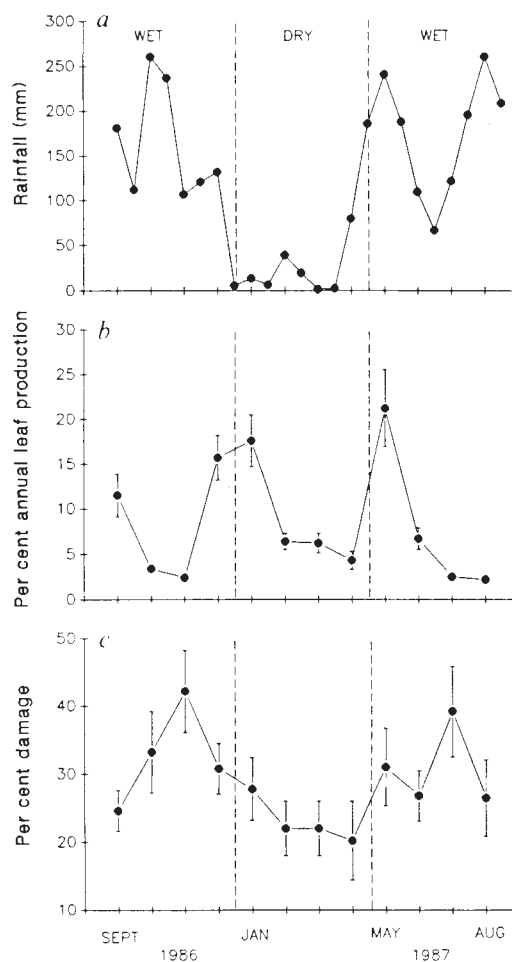


Fig. 1 Rainfall and annual patterns of leaf production and herbivory of saplings (1-2 m) of canopy and understory species in a tropical understory community of Barro Colorado Island, Panama. *a*, Rainfall values (mm) are the semi-monthly totals. Leaf production and herbivory data were collected during the first four days of each month and the mean $\pm$ s.e. values for 32 common species of woody plants are presented. *b*, The monthly proportion of annual leaf production for the 32 species. *c*, Monthly damage values are the average of mean values for all species that produced leaves in that month and were measured 30 days later when leaves were expanded. Both leaf area and damage were measured using a plastic grid. The total potential leaf area was estimated to 1 cm² and damage was estimated to 0.1 cm². This method estimates the potential photosynthetic area lost, but it does not account for the actual area eaten by herbivores (that is, holes created while a leaf is expanding will grow²³). Species censused were: *Acalypha diversifolia*, *Alseis blackiana*, *Beilschmiedia pendula*, *Capparis frondosa*, *Cousarea curvigemma*, *Desmopsis panamensis*, *Eugenia oerstedeana*, *Faramea occidentalis*, *Guarea guidonia*, *Guarea* sp. ('hairy'), *Hirtella triandra*, *Hybanthus prunifolius*, *Lacistema aggregatum*, *Mouriri myrtilloides*, *Ouratea lucens*, *Poulsenia armata*, *Pouteria unilocularis*, *Piper cordulatum*, *Prioria copaifera*, *Protium panamenses*, *P. tenuifolium*, *Psychotria horizontalis*, *Quararibea asterolepis*, *Rheedia edulis*, *Rinorea sylvatica*, *Sorocea affinis*, *Swartzia simplex*, *Tachigalia versicolor*, *Talisia princeps*, *Tetragastris panamensis*, *Trichilia tuberculata* (*T. cipo* in Croat¹⁵), *Virola sebifera*. Species nomenclature follows Croat¹⁵.

herbivores phenologically and will be defended more frequently by extrafloral nectaries.

To investigate the mechanisms of phenological defence, the timing of leaf production and damage to expanding leaves were monitored in shade-tolerant understory plants of Barro Colorado Island, Panama. The forest on this island is classified as tropical moist forest^{15,16}, but a distinct dry season of approxi-

mately four months occurs between mid December and mid April (Fig. 1a). Thirty two of the most common shade-tolerant woody species (20 individuals per species) were censused monthly for new leaf production from September 1986 to August 1987. I counted all expanding leaves and marked and remeasured a subset the following month to determine the amount of damage that occurred during the period that leaves were expanding (per cent leaf area damaged per month). From these data I calculated the proportion of the annual leaf production and damage levels to young leaves for the individual, species, and the community (all species combined) for each month.

The pattern of leaf production for the 32 species shows distinct peaks at the beginning of the wet season (May), during the wet/dry season transition (December/January) and a smaller peak in the middle of the wet season (September); together these peaks account for 66% of the annual leaf production (Fig. 1b). Rates of leaf production did not differ significantly between wet and dry season months (dry, 8.6%; wet, 8.2%; $T = 0.3$, $P > 0.7$). In four of the eight wet season months leaf production was less than 3.5% of the annual production for the community. Although plants can be water stressed during the dry season¹⁷ the three driest months (February, March and April) accounted for 15% of the annual leaf production and their monthly levels exceed those of the four lowest wet season months.

On average, 29% of the potential leaf area is lost during the first month of a leaf's lifetime. Damage to expanding leaves ranged from 42% in November to 20% in April (Fig. 1c). For the community, damage during the first month was significantly less for leaves produced during the dry season than leaves produced during the wet season (dry, 23%; wet, 32%; $T = 2.7$, $P < 0.01$). Therefore, 34% of the annual leaf production occurs during the dry season when insect abundance tends to be lowest¹⁸ and damage to expanding leaves was the lowest. These results strongly suggest that levels of damage to expanding leaves are greatly influenced by the seasonality of leaf production.

Although damage levels are lowest during the dry season, leaf production is also limited by water stress and many species produce leaves synchronously at the beginning of the wet season or during the transition from the wet to the dry season. To determine if synchronous leaf production also reduces damage to expanding leaves, I compared the annual average level of damage to each species with the damage level for the species' month of highest leaf production. Damage levels in the peak month of leaf production were lower than the annual average for 23 of the 32 species (Wilcoxon's Sign Rank Test, $T = 99$, $P < 0.005$); therefore, individuals producing leaves in synchrony with conspecifics experienced less damage.

Previous work has identified abiotic factors as the major determinants of plant phenology^{1,19-21}. Clearly, abiotic factors (for example, day length, temperature, humidity and rainfall) can act both as cues and as ultimate selective pressures on phenology. On Barro Colorado Island, higher light levels in the early dry season could facilitate leaf production in understory saplings, and water stress during the end of the dry season may limit growth. This study has shown that in addition to possible abiotic factors, herbivores appear to represent an important biotic constraint on leaf phenological patterns for shade-tolerant understory plants in this tropical moist forest. A review of flowering and fruiting phenology studies also suggests that phenological patterns can be partially explained by the selective pressures of biotic factors (for example, pollinators and predators)²². Although abiotic factors are likely to act as the proximate cues, a broader definition of phenology, including the possible influence of both abiotic and biotic factors as the ultimate selective pressures, would help in trying to understand the complex patterns of plant life-cycle events.

I thank Diana Valencia and Jillian Gregg for assistance with field work, Phyllis Coley, Diane Davidson, Sharon Emerson, Monica Geber, Tom Kursar and Jess Zimmerman for comments

on the manuscript, Mark McGinley for help with the figure, and Phyllis Coley, Egbert Leigh and Joe Wright for their advice. This work was supported by a Smithsonian Predoctoral fellowship.

Received 26 September; accepted 25 October 1988

1. Larcher, W. *Physiological Plant Ecology* (Springer, Berlin, 1980).
2. Richards, P. W. *The Tropical Rain Forest* (Cambridge Univ. Press, Cambridge, 1952).
3. Coley, P. D., Bryant, J. P. & Chapin, III, F. S. *Science* **230**, 895-899 (1985).
4. Rosenthal, G. A. & Janzen, D. J. *Herbivores: Their Interaction with Secondary Plant Metabolites* (Academic, New York, 1979).
5. Crawley, M. J. *Herbivory: the Dynamics of Animal-Plant Interactions* (Univ. California Press, Berkeley, 1983).
6. Coley, P. D. *Ecol. Monogr.* **53**, 209-223 (1983).
7. Feeny, P. *Rev. Adv. Phytochem.* **10**, 1-40 (1976).
8. McKey, D. in *Herbivores: Their Interaction with Secondary Plant Metabolites* (eds Rosenthal, G. A. & Janzen, D. H.) 55-133 (Academic, New York, 1979).
9. Mattson, W. J. Jr. *Rev. Ecol. Syst.* **11**, 119-161 (1980).
10. Scriber, J. M. & Slansky, F. Jr. *Rev. Ent.* **26**, 183-211 (1981).
11. Lieberman, D. & Lieberman, M. *Biotropica* **16**, 193-201 (1984).
12. Crawley, M. J. in *Plant Ecology* (ed. Crawley, M. J.) 253-290 (Blackwell, Oxford, 1986).
13. McKey, D. in *Proc. XIV Int. Botanical Congress* (eds Greuter, W. & Zimmer, B.) (Koeltz, Königstein, 1988).
14. Rockwood, L. L. *Ecology* **55**, 142-148 (1974).
15. Croat, T. B. *Flora of Barro Colorado Island* (Stanford Press, Stanford, 1978).
16. Leigh, E. G. Jr., Rand, A. S. & Windsor, D. M. *The Ecology of a Tropical Forest: Seasonal Rhythms and Long-term Changes* (Smithsonian Inst. Press, Washington, 1982).
17. Rundel, P. W. & Becker, P. F. *Rev. Biol. Trop.* **35** (Suppl. 1), 71-84 (1987).
18. Wolda, H. *J. anim. Ecol.* **47**, 369-381 (1978).
19. Borcher, R. *Ecology* **61**, 1065-1074 (1984).
20. Reich, P. B. & Borcher, R. *J. Ecol.* **72**, 61-74 (1984).
21. Barbour, M. G., Burk, J. H. & Pitts, W. D. *Terrestrial Plant Ecology* (Benjamin/Cummings, Menlo Park, 1987).
22. Rathcke, B. & Lacey, E. P. A. *Rev. Ecol. Syst.* **16**, 179-214 (1985).
23. Coley, P. D. *Nature* **284**, 545-546 (1980).

Incubation periods for paediatric AIDS patients

I. Auger*, P. Thomas†, V. De Gruttola‡, D. Morse§, D. Moore‡, R. Williams†, B. Truman§ & C. E. Lawrence*

* Wadsworth Center for Laboratories and Research, New York State Department of Health, P.O. Box 509, Albany, New York 12201, USA

† AIDS Surveillance, New York City Department of Health, Box 44, 125 Worth St, New York, New York 10013, USA

‡ Department of Biostatistics, Harvard School of Public Health, 677 Huntington Ave, Boston, Massachusetts 02115, USA

§ Bureau of Communicable Disease Control, New York State Department of Health, Albany, New York 12237, USA

A recent seroprevalence study of newborns indicates that one in 62 children born in New York City has antibodies to the human immunodeficiency virus (HIV)¹. The distribution of incubation periods for paediatric patients is needed to estimate future AIDS case loads from these seroprevalence data. Current estimates of incubation periods for paediatric patients are based on limited data²⁻⁵. We use parametric^{5,6} and non-parametric^{7,8} methods to analyse incubation periods for 215 paediatric patients with AIDS whose only known route of infection is maternal. We conclude that incubation periods are longer than previously reported⁹; that there is a distinct knee in the incubation period distribution at seven months which suggests two risk populations; and that there is an increase in incidence which is consistent with exponential growth.

Data were taken from paediatric AIDS patient reports in the New York City and New York State AIDS case registries¹⁰ which include all New York State residents age 12 years or less, reported through December of 1987 and diagnosed before August 1, 1987 (Table 1). AIDS diagnosis met the 1985 Centre for Disease Control definition¹¹. Of 238 patients, 23 individuals whose mode of transmission was either undetermined or through blood or blood products were removed.

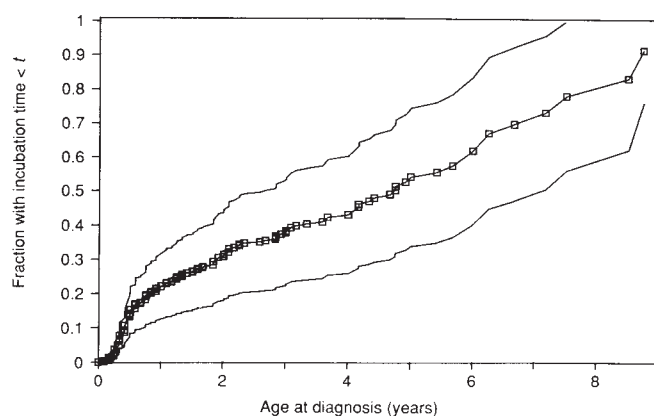


Fig. 1 Paediatric AIDS incubation times (cumulative distribution function). Non-parametric maximum likelihood estimates of the conditional distribution of incubation intervals (squares)⁷. Let all time be relative to the earliest birth, that is, February 1977. As data are available only for children who were infected and developed AIDS in a chronological time interval, realization of the infection-time and incubation-time distributions are right truncated. The estimates are obtained by a modified Kaplan-Meier product limit procedure that allows delayed entry into risk sets⁷. Entry into risk sets occurs at time of infection and failure at time 10.4 - t where t is the observed incubation time. This transformation to reverse time makes the problem of right truncation into one of delayed entry. Approximate standard error limits are obtained using the method described⁷ (lines). We have assumed that reporting was equally effective throughout for all years before August 1987. Without parametric assumptions, these data provide information only on incubation times conditional on developing AIDS in the observational interval of 10.4 years. No information is available beyond 10.4 years, or on the proportion ever to develop AIDS.

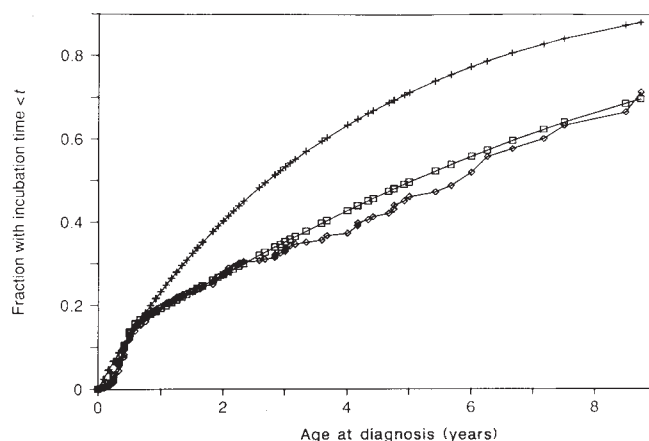


Fig. 2 Paediatric AIDS incubation times (cumulative distribution function). Estimated incubation period distributions are shown as the proportion of patients diagnosed up to a given age. Parametric estimates assume a mixture of Weibull distributions $f(t; p, \lambda_1, \gamma_1, \lambda_2, \gamma_2) = p\lambda_1\gamma_1 t^{\gamma_1-1} \exp(-\lambda_1 t^{\gamma_1}) + (1-p)\lambda_2\gamma_2 t^{\gamma_2-1} \exp(-\lambda_2 t^{\gamma_2})$, where $p, \lambda_1, \gamma_1, \lambda_2, \gamma_2$ are parameters (squares). The parameter p is the proportion of patients in the first sub-population, and t is incubation time in months. Let $\Theta = (p, \lambda_1, \gamma_1, \lambda_2, \gamma_2)$. In a similar way to that previously described⁸, the left endpoint of the truncation interval TL_i is 0 if birth occurs after February, 1979 or the number of months between birth and February 1979 (date of first diagnosis); and the right endpoint TR_i is time from birth to August 1987. Let $T_i = (TL_i, TR_i)$ and $F(T_i; \Theta)$ be $\int_{TL_i}^{TR_i} f(t; \Theta) dt$. Then, the likelihood of the observations is:

$$\prod_{i=1}^n f(T_i; \Theta) / F(T_i; \Theta).$$

Parameter estimates, obtained by maximizing this likelihood are $p = 0.12$, $\lambda_1 = 0.00342$, $\gamma_1 = 3.54$, $\lambda_2 = 0.00479$, $\gamma_2 = 1.16$ and a corresponding log likelihood of -702.75. Non-parametric estimates are maximum likelihood estimates⁸ (diamonds), and are multiplied by 0.76 to yield a comparable scale to the parametric analysis. For data that are only right truncated, other methods^{7,8} give the same results. For comparison, we also display parametric estimates with a single Weibull (+). This Weibull follows the same mathematical form as the mixture with parameters $p = 1.0$, $\lambda_1 = 0.0243$ and $\gamma_1 = 0.9615$ and a log likelihood of -738.61. Although a Weibull distribution was assumed here, previous studies suggest that mixtures of other distributions will also fit the data well². The possibility that some early patients were not reported cannot be excluded. Both the parametric⁶ and non-parametric⁸ methods used for this figure consider such potential losses under the assumption that reporting was completely ineffective in the early years of the epidemic, that is, before the date of diagnosis of the first case (February 1979). Although the parametric analysis finds that 76% of those who will ever come to diagnosis will do so in 10.4 years, this parameter and others such as the mean are sensitive to the parametric assumptions about the unobserved right tail of this distribution.

The times of infection are known only for patients who have already developed AIDS, hence patients with short incubation periods are more likely to be observed. Consequently, the frequency distribution of incubation periods, as directly observed, will be misleading in that they will underestimate the incubation periods. Here, we used non-parametric methods, with no assumptions about mathematical form, to reveal the pattern of the incubation distribution^{7,8}. Parametric methods are used to efficiently summarize the data from these distributions^{5,6}. Both methods are shown to illustrate the agreement of the assumed models with the non-parametric description.

The main results for those developing AIDS in 10.4 years are: (1) non-parametric analysis indicates that approximately 20% develop in the first year of life; (2) the remainder develop at a nearly constant rate of 8% per year, reaching the median at 4.8

Table 1 The observed distribution of incubation time by diagnosis year

Age at diagnosis	Year of diagnosis								
	1979	1980	1981	1982	1983	1984	1985	1986	1987
0-1	3	0	1	1	20	20	27	42	20
1-2	2	0	1	2	4	8	2	7	6
2-3		0	0	1	3	3	1	3	7
3-4			0	1	0	0	1	4	2
4-5				0	4	1	5	1	1
5-6					1	0	0	4	0
6-7						2	0	0	0
7-8							1	1	0
8-9								1	1

The observed numbers have been grouped into years, although they were treated by month in the analysis. The year 1987 extends from January to July. Among the mothers of the 215 patients, 150 were reported as intravenous drug users, 40 were at risk as a result of heterosexual contact with males in known high-risk groups, 17 were Haitian, six had AIDS and two were infected through blood or blood products.

Dates of HIV infection are available only for certain subpopulations. In previous studies for example, when no other risks were known, the date of transfusion was the assumed date of infection. Similarly, we assume that date of birth is the date of infection for children whose known risk is maternal transmission.

The data have limitations: although all case reports have been confirmed and the data in the records verified with hospital records, some reports may have incomplete information. There is some delay between diagnosis and the report. Some patients, especially during the early years of the epidemic, may have been assigned diagnoses other than AIDS.

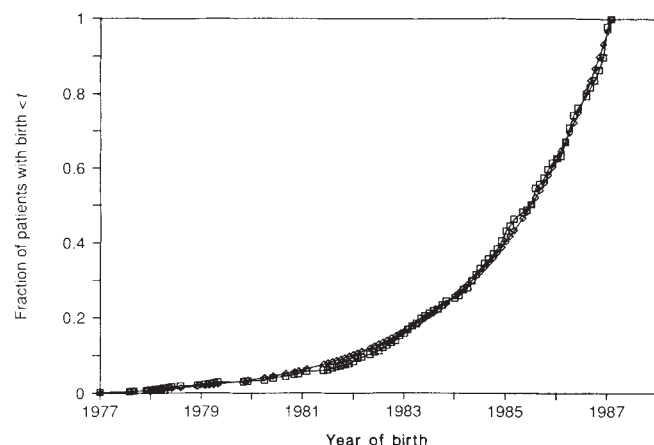


Fig. 3 Infection curves for paediatric AIDS. Estimated epidemic distributions are shown with time starting at February 1977. The right truncation time is defined as 10.4 – birth time. Non-parametric estimates (squares) are obtained by Turnbull's method⁸. Parametric estimates (diamonds) are obtained by a method similar to that of Lui and colleagues⁶. A density of the form $I(s) = \alpha \times \exp(\alpha s)$ is assumed. Estimate of the parameter α is 0.0356 (s is months since February 1977). This corresponds to a doubling time of 1.6 years which is comparable with the previously reported doubling time for transfused children of 1.3 years².

years. There is a substantially higher proportion of patients with long incubation periods than previously suspected (Fig. 1). Because of the relatively small number of patients at risk for long intervals, standard error limits of all estimates are large even with the 215 patients used here (Fig. 1).

There is an abrupt knee in the distribution at about seven months, which can be described by a mixture of distributions, suggesting the existence of two risk populations. The first distribution, including 12% + 7% (mean $\pm 2 \times$ standard error) of the patients, has very short incubation periods, median 4.1 months (Fig. 2). Because a sizeable proportion of infected infants develop AIDS within a year, their numbers may reflect changes in HIV seroprevalence in child-bearing women sooner than the number of AIDS cases amongst women does¹². The second distribution has a median incubation period of 6.1 years, comparable to that previously reported for adults of 8.0 years². The median of the second distribution may be altered as it becomes possible to observe patients with longer incubation periods, as occurred in the analysis of transfusion-related AIDS⁷.

Exponential growth in maternally infected patients fits the data better than a linear growth model (Fig. 3). Other growth models such as a cubic model in time show reasonable agreement with the non-parametric curve. The continued increase in new infections, shown in Fig. 3, seems to be contrary to other indications that the rate of increase in new HIV infections may be slowing¹³. The data used in this analysis do not, however, provide a measure that is sensitive to any changes in the shape of the curve that may have occurred recently. Furthermore, as transmission from mother to child may increase with the length of time that the mother has been infected, the curve presented here may not accurately reflect the pattern of infection in mothers. We caution against over-interpretation of this data, particularly those for the most recent years.

Incubation periods for children infected by other routes are not necessarily the same as for maternally infected children. Also, it is not clear that these results are applicable in the developing world. As the wide standard error limits clearly indicate, it will be important to re-examine these issues as more data become available. The long incubation periods reported here strongly suggest that there is a large population of children who are infected but not yet diagnosed.

Note added in proof: the New York State newborn HIV seroprevalence study has identified 1,676 seropositive newborns

among 255,521 births that occurred between 23 November 1987 and 31 October 1988. As about 40% of these are expected to be injected, this supports our suggestion that there is a large population of HIV-infected children who have not yet been diagnosed.

Received 7 June; accepted 4 November 1988.

1. *Aids in New York State* (New York State Department of Health, 1988).
2. Medley, G. F., Anderson, R. M., Cox, D. R. & Billard, L. *Nature* **328**, 719–721 (1987).
3. Medley, G. F., Billard, L., Cox, D. R. & Anderson, R. M. *Proc. R. Soc. Lond. B* **233**, 367–377 (1988).
4. Kalbfleisch, J. D. & Lawless, J. F. *Nature* **333**, 504–505 (1988).
5. Lui, K., Peterman, T. A., Lawrence, D. N. & Allen, J. M. *Statistics in Medicine* **7**, 395–401 (1988).
6. Lui, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3051–3055 (1986).
7. Lagakos, S., Barraj, L. M. & De Gruttola, V. *Biometrika* **75**, 515–523 (1988).
8. Turnbull, B. W. *J. R. Statist. Soc. B* **38**, 290–295 (1976).
9. Rogers, M. F. *et al. Pediatrics* **79**, 1008–1014 (1987).
10. *AIDS Reporting System*, (Center for Disease Control).
11. *Mortality and Morbidity Weekly Report* (Center for Disease Control) **34**, 373–375 (1985).
12. Thomas, P., O'Donnell, R., Williams, R. & Chiasson, M. A. *New Engl. J. Med.* **319**, 374 (1988).
13. *Mortality and Morbidity Weekly Report* (Center for Disease Control) **36**, (suppl S-6), 1–48 (1987).

Localization of an ataxia-telangiectasia gene to chromosome 11q22–23

Richard A. Gatti¹, Izzet Berkel², Elena Boder³, Gary Braedt⁴, Patrick Charmley^{1,5}, Patrick Concannon⁴, Fugen Ersoy⁶, Tatiana Foroud⁷, Nicholas G. J. Jaspers⁸, Kenneth Lange⁷, G. Mark Lathrop⁹, Mark Leppert⁹, Yusuke Nakamura⁹, Peter O'Connell⁹, Malcolm Paterson¹⁰, Winston Salser¹¹, Ozden Sanal^{1,6}, Jack Silver¹², Robert S. Sparkes^{3,13}, Ellen Susi¹, Daniel E. Weeks⁷, Shan Wei^{1,11}, Ray White⁹ & Freda Yoder¹⁴

Ataxia-telangiectasia (AT) is a human autosomal recessive disorder of childhood^{1,2} characterized by: (1) progressive cerebellar ataxia with degeneration of Purkinje cells; (2) hypersensitivity of fibroblasts and lymphocytes to ionizing radiation³; (3) a 61-fold and 184-fold increased cancer incidence in white and black patients, respectively⁴; (4) non-random chromosomal rearrangements in lymphocytes; (5) thymic hypoplasia with cellular and humoral (IgA and IgG2) immunodeficiencies; (6) elevated serum level of alpha-fetoprotein; (7) premature ageing; and (8) endocrine disorders, such as insulin-resistant diabetes mellitus. A DNA processing or repair protein is the suspected common denominator in this pathology⁵. Heterozygotes are generally healthy; however, the sensitivity of their cultured cells to ionizing radiation is intermediate between normal individuals and that of affected homozygotes⁶. Furthermore, heterozygous females are at an increased risk of breast cancer^{7,8}. These findings, when coupled with an estimated carrier frequency of 0.5–5.0%, suggest that (1) as many as one in five women with breast cancer may carry the AT gene⁷ and that (2) the increased radiation sensitivity of AT heterozygotes may be causing radiation therapists to reduce the doses of radiation used for treating cancer in all patients¹⁰. To identify the genetic defect responsible for this multifaceted disorder, and to provide effective

¹ Department of Pathology, UCLA School of Medicine, Los Angeles, California 90024, USA.

² Children's Medical Center, Tulsa, Oklahoma, USA.

³ Department of Pediatrics, UCLA School of Medicine, Los Angeles, California, USA.

⁴ Virginia Mason Research Center, Seattle, Washington, USA.

⁵ Department of Microbiology and Immunology, UCLA School of Medicine, Los Angeles, California, USA.

⁶ Department of Pediatrics, Hacettepe University, Ankara, Turkey.

⁷ Department of Biomathematics, UCLA School of Medicine, Los Angeles, California, USA.

⁸ Department of Cell Biology and Genetics, Erasmus University, Rotterdam, Netherlands.

⁹ Howard Hughes Medical Institute, Salt Lake City, Utah, USA.

¹⁰ Cross Cancer Institute, Edmonton, Canada.

¹¹ Department of Biology, UCLA, Los Angeles, California, USA.

¹² Hospital for Joint Diseases, New York, New York, USA.

¹³ Department of Medicine, UCLA School of Medicine, Los Angeles, California, USA.

¹⁴ University of South Carolina, Charleston, South Carolina, USA.

Table 1 Lod scores for AT versus THY1 and AT versus pYNB3.12

Family	Number affected	versus THY1 (σ)						versus pYNB3.12 (θ)					
		10^{-6}	0.05	0.1	0.2	0.3	0.4	10^{-6}	0.05	0.1	0.2	0.3	0.4
Group A													
AT004	3 ^a	0.55	0.49	0.43	0.30	0.16	0.05	0.25	0.23	0.20	0.13	0.07	0.02
AT010	1	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.10	0.08	0.05	0.02	0.01
AT011	1	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.13	0.11	0.07	0.03	0.01
AT012	3 ^{abc}	1.82	1.63	1.44	1.06	0.68	0.32	1.70	1.55	1.38	1.02	0.64	0.26
AT014	1	0.12	0.10	0.08	0.05	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00
AT018	1	0.00	0.00	0.00	0.00	0.00	0.00	-0.23	-0.18	-0.14	-0.07	-0.03	-0.01
AT019	2 ^a	0.30	0.26	0.22	0.13	0.06	0.02	0.30	0.26	0.22	0.13	0.06	0.02
TAT03	3 ^{abc}	0.84	0.74	0.64	0.42	0.22	0.06	0.18	0.15	0.12	0.07	0.03	0.01
Subtotal		3.63	3.22	2.81	1.96	1.14	0.46	2.47	2.24	1.97	1.40	0.82	0.34
Non-Group A ^d													
AT001[C]	2 ^a	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.22	0.18	0.12	0.06	0.02
AT007[C]	2 ^a	0.05	0.04	0.04	0.02	0.01	0.00	-0.01	-0.01	-0.01	-0.01	0.00	0.00
ATV20[V1]	2 ^a	0.38	0.32	0.27	0.17	0.08	0.02	0.38	0.30	0.23	0.12	0.05	0.01
Subtotal		0.43	0.36	0.31	0.19	0.09	0.02	0.62	0.51	0.40	0.23	0.11	0.03
Unassigned													
AT002	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
AT006	1	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.10	0.08	0.05	0.02	0.01
AT009	2 ^a	0.00	0.00	0.00	0.00	0.00	0.00	-4.10	-0.46	-0.23	-0.06	-0.01	0.00
AT017	2 ^{bc}	-1.34	0.12	0.28	0.30	0.23	0.12	0.58	0.48	0.38	0.22	0.09	0.02
AT021	1	-4.10	-0.48	-0.25	-0.08	-0.02	0.00	-4.50	-0.77	-0.47	-0.19	-0.07	-0.02
KAT01	1 ^c	0.12	0.09	0.07	0.03	0.01	0.00	0.26	0.20	0.15	0.07	0.03	0.01
TAT01	1	0.12	0.10	0.08	0.05	0.02	0.01	0.12	0.10	0.08	0.05	0.02	0.01
TAT02	2 ^{ac}	0.21	0.16	0.12	0.06	0.03	0.01	1.44	1.30	1.17	0.88	0.59	0.29
TAT04	3 ^{ac}	0.71	0.64	0.56	0.40	0.24	0.09	0.08	0.06	0.05	0.02	0.01	0.00
TAT05	1 ^c	-0.03	-0.04	-0.05	-0.04	-0.02	0.00	0.24	0.19	0.14	0.08	0.04	0.01
TAT08	2 ^{ac}	0.10	0.08	0.06	0.03	0.01	0.01	-3.30	0.24	0.38	0.39	0.29	0.15
TAT10	3 ^a	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.48	0.42	0.28	0.14	0.04
TAT13	3 ^{ac}	0.81	0.70	0.59	0.38	0.19	0.05	1.02	0.92	0.83	0.64	0.45	0.24
TAT14	1 ^c	0.31	0.28	0.25	0.20	0.14	0.08	0.24	0.18	0.12	0.04	0.01	0.00
TAT15	1 ^c	-3.61	-0.04	0.13	0.20	0.16	0.09	-3.61	-0.08	0.06	0.10	0.07	0.04
TAT16	1 ^c	-1.11	-0.50	-0.30	-0.13	-0.05	-0.02	-1.01	-0.39	-0.21	-0.07	-0.03	-0.01
TAT17	1 ^c	0.10	0.08	0.06	0.03	0.01	0.00	0.20	0.16	0.12	0.06	0.03	0.01
TAT18	1 ^c	0.02	0.02	0.01	0.01	0.00	0.00	-0.15	-0.11	-0.08	-0.04	-0.02	-0.01
TAT19	3 ^{ac}	-5.50	-0.84	-0.40	-0.09	-0.01	0.00	0.24	0.19	0.14	0.08	0.04	0.01
TAT20	2 ^{ac}	-0.91	-0.16	0.00	0.07	0.04	-0.01	-0.92	-0.18	-0.01	0.06	0.02	-0.02
Subtotal		-14.10	0.21	1.21	1.42	0.98	0.43	-12.50	2.61	3.12	2.66	1.72	0.78
Total		-10.04	3.79	4.33	3.57	2.21	0.91	-9.41	5.36	5.49	4.29	2.65	1.15

The number of affecteds who are typed in each pedigree is given in column 2; the superscripts on these numbers indicate the presence in the pedigree of (a) affected siblings, (b) affected cousins, and (c) inbred affecteds. The pedigrees were of varied ethnic backgrounds, including 1 Amish, 15 Turkish, 3 American Mormon, 2 Irish, 1 American Black, 1 Iranian Moslem, 1 English, 1 Phillipino, 1 Italian, 1 Ashkenazi Jewish, 1 Kuwaiti, 1 German, and 2 families of mixed ancestry. Diagnosis of AT was based on clinical evaluation, radiation-resistant DNA synthesis³, elevated serum alphafetoprotein and the presence of characteristic chromosomal aberrations. (d) Complementation group assignments are given in parentheses. V1 denotes group Variant 1 (ref. 12).

carrier detection, we performed a genetic linkage analysis of 31 families with AT-affected members. This has allowed us to localize a gene for AT to chromosomal region 11q22-23.

A potential complication in the genetic mapping of AT is the existence of at least four clinically indistinguishable complementation groups (A, C, D and E) of AT affecteds^{11,12}. Two affected individuals belong to different complementation groups if fused cells between their fibroblasts exhibit normal radiation sensitivity. It has generally been assumed that these complementation groups represent distinct AT genes. It is unknown, however, whether these genes are clustered on the same chromosome or whether they are dispersed throughout the genome.

To circumvent the complementation problem, a 61-member Amish group A pedigree (Fig. 1; ref. 13) was used to screen 171 genetic markers. The most promising marker in the Amish pedigree (AT012) turned out to be THY1 with a maximum lod score $\hat{Z} = 1.8$ at a recombination fraction $\theta = 0.00$. This THY1 *Msp*I restriction-fragment length polymorphism (RFLP) is located at chromosome 11q22.3 and has two codominant alleles, 8 and 9, with population frequencies in Caucasians of 0.72 and 0.28 respectively¹⁴. When data from the other four informative AT group A families were added (Table 1; refs 15-17), the maximum lod score rose to 3.63, with no observed recombinants.

The maximum lod score for all 31 families for linkage of AT to THY1 was $\hat{Z} = 4.34$ at $\theta = 0.10 \pm 0.04$. When separate male (θ_m) and female (θ_f) recombination fractions were estimated, the lod score attained a maximum of 4.64 with $\hat{\theta}_m = 0.18$ and $\hat{\theta}_f = 0.00$. The apparent preponderance of paternal recombination is not statistically significant ($P = 0.24$). It is noteworthy, however, that this region of chromosome 11 appears to be one of the anomalous regions of the human chromosome, showing excess male recombination¹⁸.

Encouraged by the linkage with THY1, we tested several other markers in this region. The most impressive lod scores for any marker were attained with the marker pYNB3.12 (ref. 19). This *Msp*I RFLP has two codominant alleles, 1 and 2, with population frequencies of 0.5. The maximum lod score for pYNB3.12 was $\hat{Z} = 5.58$ at $\theta = 0.08 \pm 0.03$. With separate θ_m and θ_f , $\hat{Z} = 5.97$ at $\hat{\theta}_m = 0.18$ and $\hat{\theta}_f = 0.00$. Again, these male and female recombination fractions are not statistically different ($P = 0.18$).

The ethnic backgrounds and the pedigree structures of the 30 additional AT families are briefly described in Table 1. The family data were analysed using the computer program MENDEL²⁰. The inbreeding in all the pedigrees was fully accounted for by MENDEL, with the exception of Amish pedigree AT012, which has 11 recorded generations of inbreed-

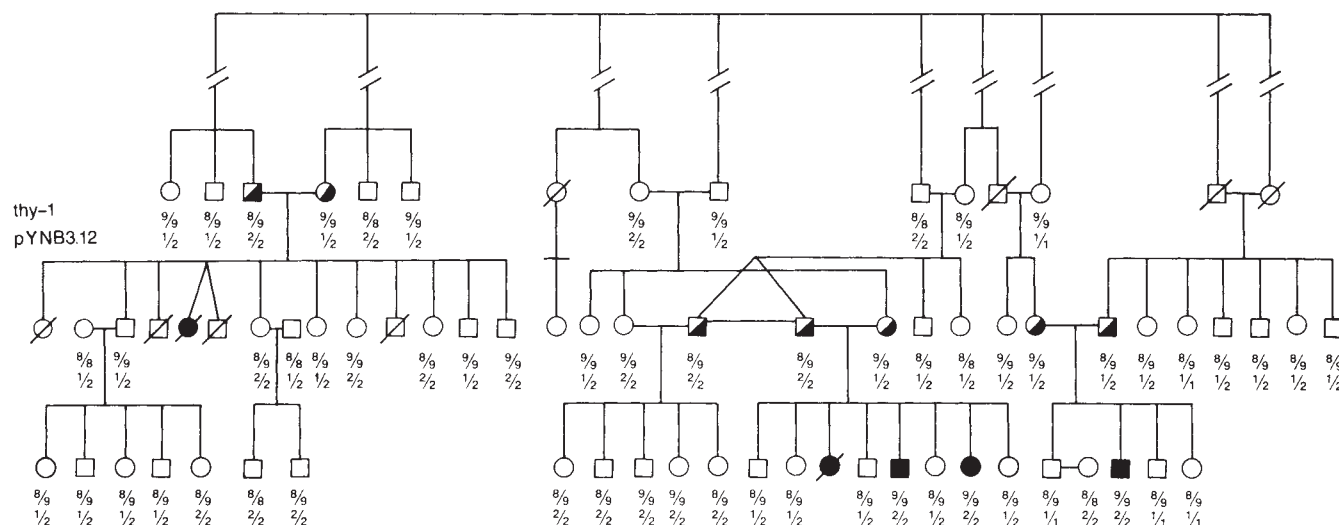


Fig. 1 Genotypes at *THY1* and *pYNB3.12* are shown for each individual in AT102, an Amish pedigree containing AT-affected members belonging to complementation group A. This American family can be traced back 11 generations to four original immigrants¹³. Lymphoblastoid cell lines were established from all living members shown here. The 171 genetic markers screened in the pedigree included blood and serum proteins, and RFLPs detected with arbitrary DNA sequences (for example, *pYNB3.12*) and cloned genes (for example, *THY1*). The latter two were determined by hybridizing DNA probes to Southern blots containing genomic DNA.

ing. The disease allele frequency was set at 0.003 (ref. 9), and we have verified that the lod scores in Table 1 are stable over a wide range of allele frequencies. All the linkage information is summarized by the lod scores; for a recessive disease like AT it is impossible to infer linkage phase and to give an unambiguous count of recombinants and nonrecombinants. A previous attempt to identify heterozygotes in the Amish pedigree, based on radiation sensitivity of fibroblasts⁶ from all 61 members, has proved to be only marginally successful and such heterozygote identification data were not used here.

Highest cumulative lod scores were attained when complementation group assignments were disregarded and all families were included in the analysis (Table 1). The unassigned pedigrees doubtless include many group A families, and this may account for the positive cumulative lod scores in the unassigned pedigrees. Furthermore, even among the non-group A pedigrees the lod scores are slightly positive for both markers. At this time, however, our data are insufficient to settle the question of genetic heterogeneity, that is, whether more than one AT gene localizes to the chromosome 11q22-23 region. More families of known complementation groups and additional closely linked markers are needed to resolve this issue. In any event, the heterogeneity question does not obscure the strong evidence in our data for linkage of at least the group A pedigrees to the chromosome 11q22-23 region.

Ordering *THY1*, *pYNB3.12* and the putative AT locus on chromosome 11 would be very helpful for future linkage studies. Our estimated recombination fraction between the two RFLPs is ~ 0.08 with $\hat{Z} = 15.97$. This estimate is based on the combined data from the 31 AT families and 40 three-generation CEPH (Centre d'Etudes des Polymorphismes Humaines) families. Our three-point linkage calculations show that the most likely order is *THY1*-*pYNB3.12*-AT, the second most likely order is AT-*THY1*-*pYNB3.12* (odds of 1:6.7 relative to the first order) and the least likely order is *THY1*-AT-*pYNB3.12* (odds of 1:601 relative to the first order).

Because of the neurological involvement and the immunodeficiencies seen in AT patients, the *THY1* marker was originally used for linkage analysis as it was a possible 'candidate gene'. *THY1* is known to be expressed in brain, thymus, fibroblasts and vascular endothelium²¹. In mice, it is expressed on the earliest defined haematopoietic stem cell²². *THY1* is also a member of the immunoglobulin supergene family²² and *THY1*-mediated T-cell activation of lymphocytes requires coexpression

of CD3 δ of the T-cell receptor (CD3/TCR) complex²⁴. *THY1* is also believed to function as an adhesion molecule in synaptogenesis²⁵. Although it would be unusually fortuitous, it is still possible that *THY1* could be the AT gene for group A.

Other known genes in the *THY1* region include those encoding apolipoproteins (A1/C3/A4), the neural cell adhesion molecule (N-CAM) and the CD3-T-cell receptor molecules (CD3 δ , CD3 γ and CD3 ϵ). It is interesting that monoclonal antibodies against CD3 react strongly with Purkinje cell neurons²⁶. The 11q22-23 region also exhibits breakpoints for translocations in certain forms of non-lymphoid leukaemias^{27,28}. Finally, the *ets-1* oncogene for avian erythroleukaemia²⁹ and two constitutional fragile sites³⁰ localize here.

In mice, a region of chromosome 9 is syntenic with human 11q22-23. This region spans at least 8 centimorgans from the gene encoding esterase A4 to the gene encoding N-CAM, with the *THY1* locus approximately in the middle³¹. The insulin-dependent diabetes susceptibility-2 locus lies centromeric to this region. Prochazka *et al.*³² have reported a linkage of a serologically detected polymorphism of *THY1* to diabetes in non-obese diabetic mice. In view of the increased frequency of diabetes in AT homozygotes (4% of deceased white patients) and in related family members (estimated relative risk of 3.8) (ref. 33), it is possible that the region of synteny between human chromosome 11q22-23 and mouse chromosome 9 may extend even closer to the centromere. Lastly, two mouse genes related to Purkinje cell degeneration also map to chromosome 9: *staggerer* and *Snell's waltzer*³¹. But both of these genes lie distal to the region of synteny with human chromosome 11q22-23. Localization of the AT group A gene within such a large region of synteny in the mouse allows its localization in other mammalian genomes; this may help to identify an animal model for ataxia telangiectasia.

Reliable genetic markers for identification of AT carriers will allow genetic counselling of families and investigation of whether AT carriers worldwide comprise a significant proportion of (1) breast cancer patients⁷, (2) all cancer patients and (3) patients who show excessive sensitivity to conventional doses of radiation therapy. Definitive mapping of AT is the first step towards understanding the complex interactions of apparently unrelated physiological systems, such as Purkinje cell degeneration, immunodeficiency and radiosensitivity.

This work was supported by the US Department of Energy, the US Public Health Service, the American Cancer Society, a USPHS National Research Service Award, the Commission of

European Communities, the Ataxia-Telangiectasia Medical Research Foundation, Children's World, and the Joseph Drown Foundation. R.W. is an Investigator of the Howard Hughes Medical Institute. We thank Michael Boehnke, Richard Davis, Jean-Marc Lalouel, Michael Litt, Victor McKusick, Rahmi Pehlivanli, Roy Shaked, Pamela and George Smith and the many AT families studied.

Received 25 July; accepted 25 October 1988.

- Boder, E. & Sedgwick, R. P. *Pediatrics* **21**, 526-554 (1958).
- Boder, E. in *Ataxia-telangiectasia: Genetics, Neuropathology and Immunology of a Degenerative Disease of Childhood* (eds Gatti, R. A. & Swift, M.) 1-61 (Liss, New York, 1985).
- Painter, R. B. in *Ataxia-telangiectasia: Genetics, Neuropathology and Immunology of a Degenerative Disease of Childhood* (eds Gatti, R. A. & Swift, M.) 89-100 (Liss, New York, 1985).
- Morrell, D., Cromartie, E. & Swift, M. *J. natn. Cancer Inst.* **77**, 89-92 (1986).
- Hanawalt, P. & Painter, R. in *Ataxia-telangiectasia: Genetics, Neuropathology and Immunology of a Degenerative Disease of Childhood* (eds Gatti, R. A. & Swift, M.) 67-71 (Liss, New York, 1985).
- Paterson, M. C., MacFarlane, S. J., Gentner, N. E. & Smith, B. P. in *Ataxia-telangiectasia: Genetics, Neuropathology and Immunology of a Degenerative Disease of Childhood* (eds Gatti, R. A. & Swift, M.) 73-87 (Liss, New York, 1985).
- Swift, M., Reitnauer, P. J., Morrell, D. & Chase, C. L. *New Engl. J. Med.* **316**, 1289-1294 (1987).
- Pippard, E. C., Hall, A. J., Barker, D. J. & Bridges, B. A. *Cancer Res.* **48**, 2929-2933 (1988).
- Swift, M. *et al. Am. J. hum. Genet.* **39**, 573-583 (1986).
- Norman, A., Fagan, R. A. & Chan, S. L. *Am. J. Oncol.* **11**, 84-88 (1988).
- Jaspers, N. G. J., Painter, R. B., Paterson, M. C., Kidson, C. & Inoue, T. in *Ataxia-telangiectasia: Genetics, Neuropathology and Immunology of a Degenerative Disease of Childhood* (eds Gatti, R. A. & Swift, M.) 147-162 (Liss, New York, 1985).
- Jaspers, N. G. J., Gatti, R. A., Baan, C., Linssen, P. C. M. L. & Bootsma, D. *Cytogenet. Cell Genet.* (in press).
- Çinter, D. N. & Tallapragada, R. in *Medical Genetic Studies of the Amish* (ed. McKusick, V.) 144-145 (Johns Hopkins Univ. Press, Baltimore, 1978).
- Gatti, R. A. *et al. Hum. Immun.* **22**, 145-150 (1988).
- Ott, J. *Analysis of Human Genetic Linkage* (Johns Hopkins Univ. Press, Baltimore, 1985).
- Hodge, S. E., Anderson, C. E., Neiswanger, K., Sparkes, R. S. & Rimoin, D. L. *Am. J. hum. Genet.* **35**, 1139-1155 (1988).
- Smith, C. A. B. *Ann. hum. Genet.* **27**, 175-182 (1963).
- Donis-Keller, H. *et al. Cell* **51**, 319-337 (1987).
- Nakamura, Y. *et al. Science* **235**, 1616-1622 (1987).
- Lange, K., Boehnke, M. & Weeks, D. E. *Programs for Pedigree Analysis* Department of Biomathematics, University of California, Los Angeles (1988).
- Gordon, J. W. *et al. Cell* **50**, 445-454 (1987).
- Spangrude, G., Heimfeld, S. & Weissman, I. L. *Science* **241**, 58-62 (1988).
- Williams, A. F. *Nature* **314**, 579-580 (1985).
- Gunter, K. C. *et al. Nature* **362**, 505-507 (1987).
- Acton, R. T., Addis, J., Carl, G. F., McClain, L. D. & Bridges, W. F. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3283-3287 (1978).
- Garson, J. A., Beverly, P. C. L., Coakham, H. B. & Harper, E. I. *Nature* **298**, 375-377 (1982).
- Kameko, Y. *et al. Blood* **67**, 484-491 (1986).
- Hagemeyer, A., Hahden, K., Sizoo, W. & Abels, J. *Cancer Genet. Cytogenet.* **5**, 95-105 (1982).
- Rovigatti, U., Watson, D. K. & Yunis, J. J. *Science* **232**, 398-400 (1986).
- Yunis, J. J., Brunning, R. D., Howe, R. B. & Lobell, M. *New Engl. J. Med.* **311**, 812-818 (1984).
- Davison, M. T. & Roderick, T. H. *Mouse News Letter* **81**, 1-8 (1988).
- Prochazka, M., Leiter, E. H., Serreze, D. V. & Coleman, D. L. *Science* **237**, 286-289 (1987).
- Swift, M. in *Ataxia-telangiectasia: Genetics, Neuropathology and Immunology of a Degenerative Disease of Childhood* (eds Gatti, R. A. & Swift, M.) 133-146 (Liss, New York, 1985).

Analysis of specificity for antigen, Mls, and allogeneic MHC by transfer of T-cell receptor α - and β -chain genes

Jonathan Kaye & Stephen M. Hedrick

Department of Biology and Cancer Center, University of California at San Diego, La Jolla, California 92093, USA

The majority of peripheral T lymphocytes bear cell-surface antigen receptors comprised of a disulphide-linked $\alpha\beta$ dimer¹. In an immune response, this receptor endows T cells with specificities for foreign antigenic protein fragments bound to cell surface glycoproteins encoded in the major histocompatibility complex (MHC)²⁻⁵. At a high frequency (>1%), the same population of T lymphocytes responds to allogeneic MHC glycoproteins⁶⁻⁸, or to differences at other genetic loci termed Mls^{8,9}, in conjunction with MHC¹⁰. The $\alpha\beta$ -antigen receptor has been implicated in alloreactivity¹¹⁻¹³ and Mls reactivity¹⁴⁻¹⁸. In fact, many monoclonal T-cell lines recognize a foreign protein fragment bound to self-MHC molecules and, in addition, recognize allogeneic MHC glycoproteins¹⁹⁻²¹, an Mls-encoded determinant²²⁻²⁴, or both²⁵. For at least one T-cell clone, a monoclonal antibody directed against the $\alpha\beta$ antigen receptor has been shown to block activation induced by either antigen-bound self-MHC or by allogeneic MHC²⁶. However, it remains to be demonstrated directly that a single $\alpha\beta$ receptor can mediate antigen specificity, alloreactivity and Mls reactivity, a prerequisite to understanding the structural basis of these high-frequency cross-reactivities. To address this issue we have performed transfers of receptor chain genes from a multiple-reactive T-cell clone into an unrelated host T lymphocyte. We now demonstrate definitively that the genes encoding a single $\alpha\beta$ -receptor chain pair can transfer the recognition of self-MHC molecules complexed with fragments of antigen, allogeneic MHC molecules, and an Mls-encoded determinant (presumably in conjunction with MHC). In this case the transfer of antigen specificity and alloreactivity requires a specific $\alpha\beta$ -receptor chain combination, whereas Mls reactivity can be transferred with the β -chain gene alone into a recipient expressing a randomly selected α -chain.

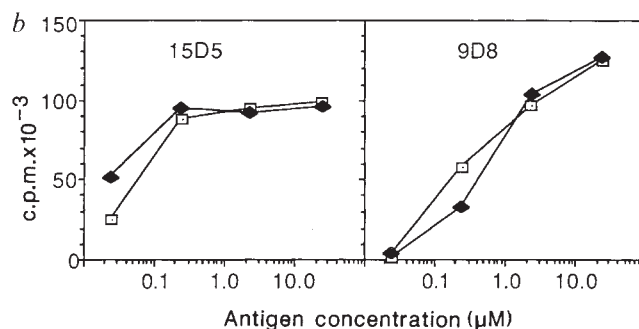
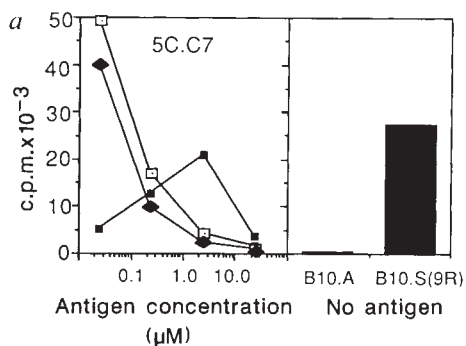
The alloreactive T-cell clone 5C.C7 (C7) was derived from a B10.A mouse immunized with pigeon cytochrome *c*. The sequences of the productively rearranged α - and β -chain receptor

genes from these cloned T cells have been reported previously²⁷. C7 has rearranged $V_{\beta}3$ to $J_{\beta}1.2$ and $V_{\alpha}11.1$ to $J_{\alpha}C7$. Clone C7 is specifically activated by a C-terminal peptide of pigeon or moth cytochrome *c* presented by E^k-bearing antigen-presenting cells (APC) (Fig. 1a). The proliferative response of this clone is inhibited at high concentrations of antigen, in a similar way to that previously reported for many cytochrome *c*-specific T-cell clones²⁷. In addition, C7 recognizes A^s-bearing APC derived from B10.S(9R) mice, in the absence of exogenously added cytochrome *c* antigens (Fig. 1a).

C7 α - and β -receptor chain genes (see Fig. 1) were transfected into the murine helper T-cell clone D10.G4.1 (D10). As reported previously²⁶, D10 cells are specific for the egg-white protein conalbumin in conjunction with A^k, and are themselves alloreactive to the A^b molecule. D10 cells do not cross-react with pigeon or moth cytochrome *c* fragments in conjunction with E^k or A^k, nor do they exhibit any reactivity to A^s or Mls-2 (data not shown, but see control transfectants in Figs 2 and 3). In addition, D10 and C7 use unrelated V_{α} and V_{β} gene segments (data not shown). Of the 47 D10 transfectants tested in three experiments, 14 could be activated by pigeon cytochrome *c* and E^k. Figure 1b depicts the response of two such D10 transfectants (15D5 and 9D8) to pigeon and moth cytochrome *c* fragments in the presence of E^k bearing APC, as measured by lymphokine release. Typically for the transfectants, these two clones show a dose-response curve shifted to higher concentrations of antigen relative to the C7 donor clone. Presumably, this reflects lower numbers of receptors on the transfectants. Similar results were obtained when proliferation was used as a measure of activation (data not shown), although D10 transfectants do not show a decreased response at high concentrations of antigen. This may reflect the fact that D10 and C7 show a different profile of lymphokine secretion (data not shown), and thus may represent different types of T helper cells.

These transfectants were also tested for alloreactivity, and the results of these experiments are shown in Fig. 2a. The $\alpha\beta$ -chain transfectants 15D5 and 9D8 respond to B10.S(9R) spleen cells in the absence of antigen, similar to the response of donor T-cell clone C7. To eliminate the possibility of a cross-reaction on K^s or E^s due to D10 and C7 mixed receptor-chain combinations the specificity of the alloreactivity was mapped to a sub-region of the MHC locus. The response to A.TH, but not A.TL APC in the absence of antigen maps the alloreactivity to A^s. The A.TL APC were capable of stimulating a response in the presence of either pigeon cytochrome *c* (E^k) or conalbumin (A^k).

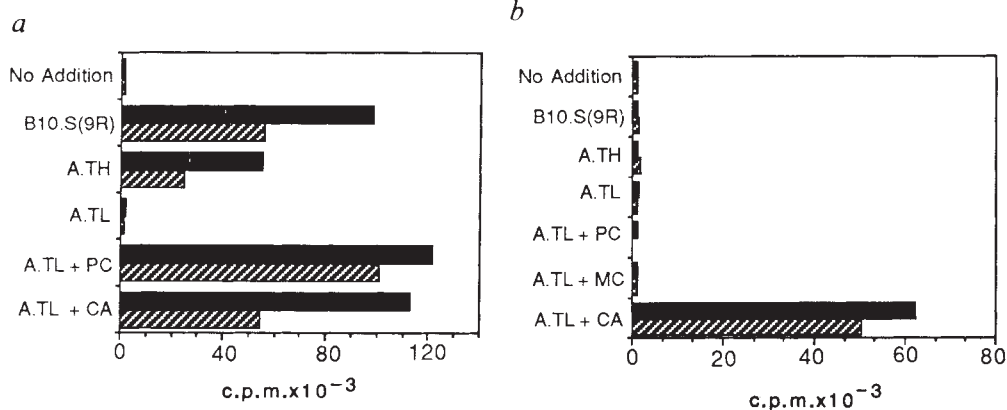
Fig. 1 Antigen specificity is transferred with C7 α - and β -chain genes. *a*, 2×10^4 C7 cells were cultured with 5×10^5 irradiated ($3,500$ rad) B10.A spleen cells in 96-well flat bottom microtitre plates in the presence of various concentrations of whole pigeon cytochrome *c* (■), or synthetically produced C-terminal peptides of



pigeon cytochrome *c* (residues 88–104) (□) or moth cytochrome *c* (residues 88–103) (◆). Alternatively, T cells were cultured with B10.A or B10.S(9R) irradiated spleen cells in the absence of cytochrome *c* (bars). After 48 h, cultures were pulsed with $1 \mu\text{Ci} [^3\text{H}]$ thymidine for 18 h, cells collected and isotope incorporation determined. *b*, D10 transfectants 15D5 and 9D8 expressing C7 α - and C7 β -receptor chain genes were cultured as described above in the presence of B10.A irradiated spleen cells and pigeon (□) or moth (◆) cytochrome *c* peptides. After 24 h in culture, supernatants were collected and assayed for IL-4 activity by addition to 5×10^3 cells of the IL-2/IL-4-dependent cell line, NK³², at a final supernatant concentration of 10%. After 24 h NK cell cultures were pulsed and processed as above. There was no detectable response of NK cells to 10% supernatants derived from 9D8 or 15D5 cells cultured in the absence of cytochrome *c* over a mean background incorporation of 1,798 c.p.m. All data presented are the mean thymidine incorporation of triplicate cultures.

Methods. A C7 α -chain cDNA lacking the polyadenylation signal was cloned into the *Bam*HI site of the expression vector pH β APr-1-neo (ref. 33). This vector contains a human β -actin promoter for cDNA expression and a neomycin-resistance gene, encoding resistance to compound G418 in mammalian cells, as a selectable marker. The C7 rearranged β -chain gene contained on a 8.6-kilobase (kb) *Hind*III fragment of genomic DNA was isolated from a J1 λ phage library. A 700-base pair (bp) *Xba*I-*Eco*RI fragment containing the immunoglobulin heavy-chain enhancer³⁴ has been cloned into the *J-C* intron of this gene. To produce transfectants, D10 cells were washed into ice-cold Dulbecco's PBS supplemented with 10 mM MgCl₂, resuspended in the same at 1×10^7 cells ml⁻¹ in the presence of 100 μg ml⁻¹ linearized DNA, and incubated on ice for 10 min. C7 α - and C7 β -gene constructs were used at a molar ratio of 1:5 respectively. The cell-DNA mixture (0.25 ml) was subjected to a single 500- μs pulse at 2.5 kV cm⁻¹ using a BTX-100 electroporator power supply. After an additional 10-min incubation on ice, cells were diluted with 1 ml of serum-free medium and incubated at room temperature for 10 min. Cells were then diluted into serum-containing medium and plated into 96-well flat-bottom microtitre plates at a concentration of 2.5×10^4 input T cells per well with the addition of 2.5×10^5 irradiated spleen cells per well and 2.5 μg ml⁻¹ concanavalin A. Medium containing G418 and recombinant human IL-2 (ref. 35; generously provided by Cetus Corp.) was added after 48 h to give final concentrations of 400 μg ml⁻¹ and 50–100 U ml⁻¹ respectively. Cells were restimulated as above after approximately 7 days and then as needed for expansion. G418-resistant clones were screened for pigeon cytochrome *c* specificity and subsequently maintained by biweekly stimulation with pigeon cytochrome *c* and H-2^k APC followed by expansion in IL-2.

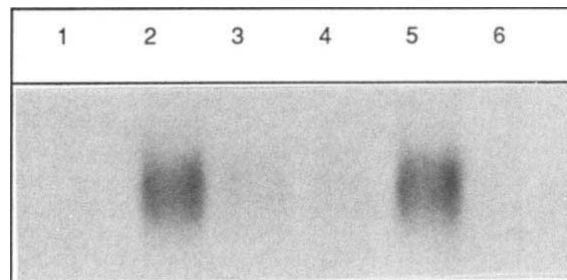
Fig. 2 Transfer of alloreactivity requires expression of both C7 α and C7 β . The response of T-cell transfectants to B10.S(9R), A.TH, or A.TL APC was determined by lymphokine secretion as described in the Fig. 1 legend. Where indicated, 2.4 μM pigeon cytochrome *c* fragment 88–104 (PC), 2.4 μM moth cytochrome *c* fragment 88–103 (MC), or 200 μg ml⁻¹ concanavalin (CA) was added to the cultures. The expressed class I and class II MHC alleles for relevant mouse strains are indicated. *a*, Response of C7 α transfectants 9D8 (black bars) and 15D5 (striped bars). *b*, Response of C7 α B10 β transfectants 4F12 (black bars) and 4D2 (striped bars). *c*, Total cytoplasmic RNA was prepared from various cells, separated by electrophoresis in formaldehyde agarose gels, transferred to nitrocellulose and probed with a nick-translated [³²P] labelled V α 11 probe. All lanes contain 12 μg RNA. Lanes 1, D10; 2, 9D8; 3, 15D5; 4, 4D2; 5, 4F12; 6, 30F3.



	K	A	E β	D
B10.S(9R)	s	s	s	d
A.TH	s	s	-	d
A.TL	s	k	k	d

Transfected genes

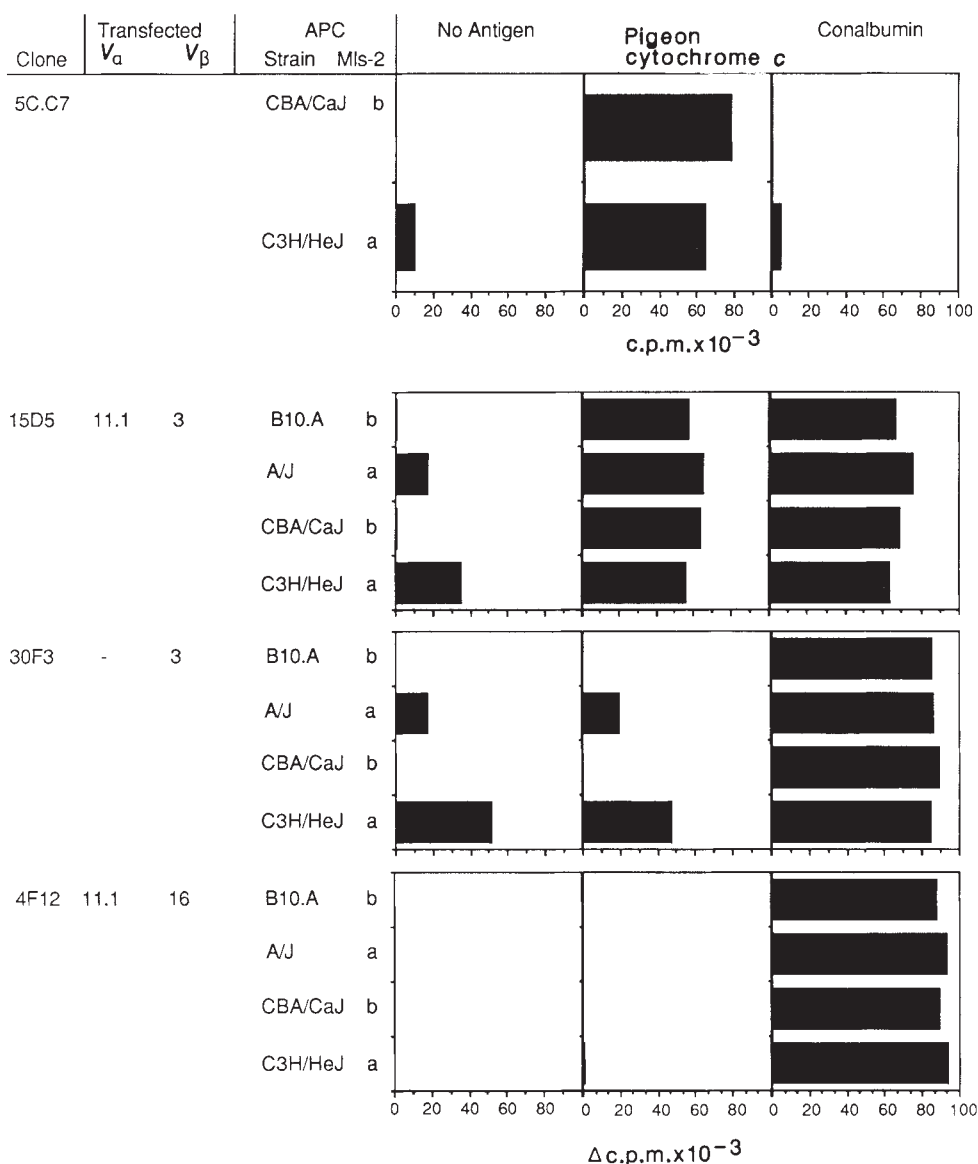
V α	-	11.1	11.1	11.1	11.1	-
V β	-	3	3	16	16	3



V α 11 probe

Methods. 4F12 and 4D2 were produced by cotransfection of D10 cells with a C7 α -cDNA construct (See Fig. 1 legend) and a β -chain cDNA also cloned into the *Bam*HI site of expression vector pH β APr-1-neo. This β -chain cDNA was derived from the pigeon cytochrome *c*-specific T-cell clone B10 which expresses V β 16-J β 1.2 (ref. 27). Transfections were performed as described in the Fig. 1 legend but with equimolar amounts of linearized α - and β -genes. 30F3 was produced similarly by co-transfection of D10 cells with a genomic C7 β clone (see Fig. 1) in fivefold molar excess over the selectable plasmid pH β APr-1-neo. G418-resistant clones were screened by RNA dot blots with appropriate V-region probes. Selected transfectants were then maintained by biweekly stimulation with concanavalin A and irradiated spleen cells.

Fig. 3 Mls reactivity is transferred with the C7 β -gene alone. The response of T-cell clones to various APC was determined by proliferation (C7) or lymphokine release (transfectants 15D5, 30F3 and 4F12) as described in Fig. 1. For the latter, the response of NK cells to 20% culture supernatants is shown. Where indicated conalbumin (200 $\mu\text{g ml}^{-1}$) or pigeon cytochrome *c* peptide (0.024 μM for C7, others 2.4 μM) was added to cultures. NK cell background proliferation in the absence of culture supernatants ranged from 986 to 2044 c.p.m. and has been subtracted from the values shown. The sources of APC were as follows; C3H/HeJ and CBA/CaJ mice were given a single intravenous injection of 150 μg of a monoclonal anti-IgD antibody (AMS9.1) which recognizes the Igh5.4 specificity present in both these strains³⁶. It has been reported previously that such treatment enhances responses to Mls-2 (ref. 31). 20–24 h after injection mice were killed, spleen cells collected and mitomycin *c*-treated. Because monoclonal antibody AMS9.1 does not react with strains B10.A or A/J, spleen cells from these strains were cultured for 24 h at 4×10^6 per ml in the presence of 50 $\mu\text{g ml}^{-1}$ lipopolysaccharide, and then treated with mitomycin *c* before use. APC from a given set of Mls-2^a and Mls-2^b strains were always treated identically and assayed on the same day. Data represent the mean [³H] thymidine incorporation of triplicate cultures.



The alloreactivity of T-cell clone C7 has been associated with the $J_\alpha C7$ gene segment which has thus far been found only in cytochrome *c*-specific T cells cross-reactive with A^S , and rearranged to $V_\alpha 11^{12,27}$. In addition, in alloreactive T-cell lines this α -chain is paired only with β -chains consisting of $V_\beta 3$ and $J_\beta 1.2$, suggesting that the alloreactivity may require a particular $\alpha\beta$ -chain combination¹³. For this reason we determined the ability of the C7 α -chain to impart A^S reactivity in the absence of the C7 β -chain. The C7 α -chain was cotransfected into D10 cells in the presence of a $V_\beta 16$ complementary DNA construct derived from the phenotypically distinct pigeon cytochrome *c*-specific T-cell clone B10²⁷. Transfectants were screened by RNA dot blots using a $V_\alpha 11$ probe. Two D10 transfectants, 4D2 and 4F12, that expressed $V_\alpha 11$ RNA levels comparable to those observed in C7 $\alpha\beta$ transfectants were selected for further analysis (Fig. 2c). It should be noted that the C7 α -chain had the potential to pair with either of two β -chains in this experiment, the endogenous β -chain and the transfected β -chain, as both transfectants do express low levels of $V_\beta 16$ (not shown). However, as shown in Fig. 2b, these transfectants were not responsive to A^S -bearing APC, nor to pigeon cytochrome *c* and E^k , although they still maintained the host-cell specificity for conalbumin and A^k . Although reagents are not available to detect cell-surface expression of the transferred α -chain in this experiment, our

lab has not found a discordance between mRNA expression and functional expression in analysing other D10 transfectants (data not shown; ref. 28). In addition, transfections have demonstrated that another $V_\alpha 11$ -bearing α -chain derived from the pigeon cytochrome *c*-reactive hybridoma 2B4 (ref. 29), can pair with the D10 β -chain, as detected by a monoclonal anti-receptor antibody (data not shown).

D10 cells transfected with the C7 β -gene alone were also not responsive to A^S (not shown) or pigeon cytochrome *c* and E^k , although the transfected β -chain gene could be detected functionally as discussed below. We conclude from these experiments that a single receptor contains sites for both antigen complexed with self-MHC and allogeneic MHC (either alone or complexed with an unknown peptide). In addition, both sites are dependent on the interaction of specific α - and β -receptor chains.

Although originally described as a single locus mapped to chromosome 1³⁰, the Mls system now seems to comprise at least two unlinked loci³¹ Mls-1 (formerly designated Mls^a) and Mls-2 (formerly designated Mls^c), each with a stimulatory allele, *a*, and a non-stimulatory allele, *b*. Lymphocytes from mice possessing a *b* allele at both loci are stimulated by Mls-disparate cells but are themselves not stimulatory.

A series of pigeon cytochrome *c* reactive T-cell clones includ-

ing C7 have recently been shown to react to Mls-2^a. We have confirmed this observation and, as shown in Fig. 3, T-cell clone C7 responds to C3H/HeJ APC (H-2^k, Mls-1^{b2a}), but not CBA/CaJ APC (H-2^k, Mls-1^{b2b}) in the absence of antigen.

On the basis of this observation, it was interesting to test whether Mls reactivity in addition to antigen/MHC and alloreactivity could be transferred with one T-cell receptor $\alpha\beta$ -chain pair. Both a transfectant expressing the C7 α - and C7 β -chains, as well as a transfectant expressing only the C7 β -chain were responsive to Ia^k-bearing APC from two Mls-1^{b2a} strains (C3H/HeJ and A/J) but did not respond to Ia^k-bearing APC from two Mls-1^{b2b} strains (CBA/CaJ and B10.A) (Fig. 3). Although mouse strains A.TH and A.TL are also Mls-2^a, the APC in these experiments (Fig. 2) were irradiated and unstimulated (compare Figs 2 and 3) which precluded detection of the Mls-2^a response. Transfectants 4F12 and 4D2 which express the C7 α -chain and no V β 3 are not Mls-reactive (Fig. 3). Although all D10 transfectants maintained specificity for conalbumin/A^k, only C7 $\alpha\beta$ -double transfectants were responsive to pigeon cytochrome c/E^k (Fig. 3). Thus, in comparison to antigen recognition or allogeneic MHC recognition, recognition of Mls-2^a places fewer structural constraints on the $\alpha\beta$ receptor. This is consistent with the recent demonstration that expression of either of two V β gene segments in a functional receptor is sufficient to confer Mls-1^a reactivity to CD4-bearing T cells^{14,15}. Although we have not formally mapped the responses of D10 transfectants to Mls-2, the correlations observed are strong evidence for gene transfer of Mls-2 specificity. Furthermore, recent reports have shown that the reactivity of V β 3-bearing T-cell clones maps to Mls-2¹⁶⁻¹⁸.

We have shown that a single receptor can cross-react with multiple ligands, in this case pigeon or moth cytochrome c fragments and E^k, A^S (and an unknown peptide?), and an Mls-2 controlled determinant (and A^k and/or E^k?). This result implies that the high frequency of allogeneic MHC reactivity and Mls reactivity results from the cross-reaction of self-MHC restricted, antigen-specific receptors with these determinants; all three specificities can be mediated by the same $\alpha\beta$ -chain pair. These results are in apparent conflict with a recent study²⁵ that used unstable T-cell hybrids to demonstrate independent segregation of antigen/MHC reactivity and alloreactivity from Mls-1^a reactivity. However, it is possible that in some cases Mls reactivity is more dependent than alloreactivity or antigen/MHC reactivity upon an additional T-cell accessory molecule needed to enhance the avidity of the T cell-APC interaction.

We thank Mei-Ling Hsu for technical assistance, Dr Louis A. Matis for discussions and critical reading of this manuscript, and Dr P. Linton for the gift of anti-IgD monoclonal antibody. J.K. was supported by a fellowship from the Cancer Research Institute/F. M. Kirby Foundation, and S.M.H. was supported by a Research Career Development Award from the National Institute of Allergy and Infectious Diseases. This work was funded by research grants from the NIH.

Note added in proof: Malissen and co-workers³⁷ have recently demonstrated that a single $\alpha\beta$ receptor-chain pair can transfer both antigen specificity and allogeneic MHC specificity.

15. MacDonald, H. R. *et al.* *Nature* **332**, 40-45 (1988).
16. Abe, R., Nacchio, M. S., Fox, B. & Hodes, R. J. **335**, 827-830 (1988).
17. Fry, A. M. & Matis, L. A. *Nature* **335**, 830-832 (1988).
18. Pullen, A., Marrack, P. & Kappler, J. *Nature* **335**, 796-801 (1988).
19. Janeway, C. A. Jr, Lerner, E. A., Conrad, P. J. & Jones, B. *Behring Inst. Mitt.* **70**, 200-209 (1982).
20. Butz, E. *et al.* *Immunogenetics* **22**, 189-192 (1985).
21. Ashwell, J. D., Chen, C. & Schwartz, R. H. *J. Immun.* **136**, 389-395 (1986).
22. Braciale, V. L. & Braciale, T. J. *J. Immun.* **127**, 859-862 (1981).
23. Katz, M. E. & Janeway, C. A. Jr *J. Immun.* **134**, 2064-2070 (1985).
24. Lynch, D. H., Gress, R. E., Needleman, B. W., Rosenberg, S. A. & Hodes, R. J. *J. Immun.* **134**, 2071-2078 (1985).
25. Webb, S., Okamoto, A. & Sprent, J. *J. Immun.* **141**, 1828-1834 (1988).
26. Kaye, J. & Janeway, C. A. Jr *J. exp. med.* **159**, 1397-1412 (1984).
27. Fink, P. J., Matis, L. A., McElligott, D. L., Bookman, M. & Hedrick, S. M. *Nature* **321**, 219-226 (1986).
28. Engel, I. & Hedrick, S. M. *Cell* **54**, 473-484 (1988).
29. Samelson, L. E., Germain, R. N. & Schwartz, R. N. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6972-6976 (1983).
30. Festenstein, H., Bishop, C. & Taylor, B. A. *Immunogenetics* **5**, 357-361 (1977).
31. Abe, R., Ryan, J. J. & Hodes, R. J. *J. exp. Med.* **165**, 1113-1129 (1987).
32. Pfeiffer, D. H., McKenzie, D. T., Swain, S. L. & Putton, R. W. *J. exp. Med.* **166**, 1464-1470 (1987).
33. Gunning, P., Leavitt, J., Muscat, G., Ng, S.-Y. & Uedes, L. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4831-4835 (1987).
34. Sen, R. & Baltimore, D. *Cell* **46**, 705-716 (1986).
35. Wang, A., La, S.-D. & Mark, D. *Science* **224**, 1431-1433 (1984).
36. Stall, A. M. & Loken, M. R. *J. Immun.* **132**, 787-795 (1984).
37. Malissen *et al.* *Cell* **55**, 49-59 (1988).

Inositol 1,4,5-trisphosphate activates a channel from smooth muscle sarcoplasmic reticulum

B. E. Ehrlich & J. Watras

Departments of Medicine and Physiology, University of Connecticut, Farmington, Connecticut 06032, USA

Inositol 1,4,5-trisphosphate (InsP₃) can initiate calcium release into the cytoplasm in a variety of cells^{1,2}. From experiments using permeabilized cells^{3,4}, membrane vesicles⁵⁻⁹, and patch-clamp techniques^{10,11}, it has been suggested that InsP₃ acts by directly opening calcium channels. Here, we show that InsP₃ induced openings of channels in planar lipid bilayers into which vesicles made from aortic muscle sarcoplasmic reticulum (SR) were incorporated. Activation of channels by InsP₃ was not observed when vesicles made from SR of cardiac or skeletal muscle were incorporated into planar lipid bilayers. The present study demonstrates for the first time unique properties of an InsP₃-gated calcium channel in sarcoplasmic reticulum vesicles from vascular smooth muscle. This InsP₃-activated channel from aortic SR differs strikingly from the calcium-gated channel of striated muscle SR in single-channel conductance and pharmacology.

Different compounds initiate calcium release from SR vesicles made from aortic and skeletal muscle. Sarcoplasmic reticulum vesicles isolated from canine aortic smooth muscle or rabbit skeletal muscle accumulated calcium in the presence of ATP. Addition of InsP₃ to aortic SR vesicles induced calcium release (Fig. 1a and ref. 7), but not from skeletal SR vesicles (Fig. 1c and ref. 7). In contrast, addition of caffeine to aortic SR vesicles did not initiate calcium release (Fig. 1b), but did induce calcium release in skeletal SR vesicles (Fig. 1c). Similar results were obtained using SR from cardiac muscle¹². Although InsP₃-induced calcium release has been reported for skeletal muscle SR^{13,14} and cardiac muscle SR^{8,15}, the presence of an InsP₃-gated calcium release pathway in striated muscle SR is controversial^{7,16-18}. Moreover, the time course of InsP₃-induced calcium release from skeletal SR is too slow for InsP₃ to be the signal in excitation-contraction coupling in skeletal muscle¹⁹.

InsP₃ activation of calcium release from aortic SR vesicles is specific. Several other inositol phosphates (inositol 1-phosphate, inositol 1,4-bisphosphate, inositol 1,3,4-trisphosphate, inositol 1,3,4,5-tetrakisphosphate; each at 20 μ M) and caffeine (1-50

Received 15 August; accepted 24 October 1988.

1. Kronenberg, M., Sin, G., Hood, L. E. & Shastri, N. A. *Rev. Immun.* **4**, 529-591 (1986).
2. Zinkernagel, R. M. & Doherty, P. C. *J. exp. Med.* **141**, 1427-1436 (1975).
3. Babbitt, B., Allen, P. M., Matsueda, G., Haber, E. & Unanue, E. *Nature* **317**, 359-361 (1985).
4. Dembic, Z. *et al.* *Nature* **320**, 232-238 (1986).
5. Buus, S., Sette, A., Colon, S. M., Miles, C. & Grey, H. M. *Science* **235**, 1353-1358 (1987).
6. Lindahl, K. F. & Wilson, D. B. *J. exp. Med.* **145**, 508-522 (1977).
7. MacDonald, H. R. *et al.* *Immun. Rev.* **51**, 93-123 (1980).
8. Miller, R. A. & Stutman, O. *J. Immun.* **128**, 2258-2264 (1982).
9. Lutz, C. T., Glasebrook, A. L. & Fitch, F. W. *Eur. J. Immun.* **11**, 726-734 (1981).
10. Janeway, C. A. Jr & Katz, M. E. *J. Immun.* **134**, 2057-2063 (1985).
11. Gabert, J. *et al.* *Cell* **50**, 545-554 (1987).
12. Sorger, S. B., Hedrick, S. M., Fink, P. J., Bookman, M. A. & Matis, L. A. *J. exp. Med.* **165**, 279-301 (1987).
13. Matis, L. A., Sorger, S. B., McElligott, D. L., Fink, P. J. & Hedrick, S. M. *Cell* **51**, 59-69 (1987).
14. Kappler, J. W., Staerz, V., White, J. & Marrack, P.-C. *Nature* **332**, 35-40 (1988).

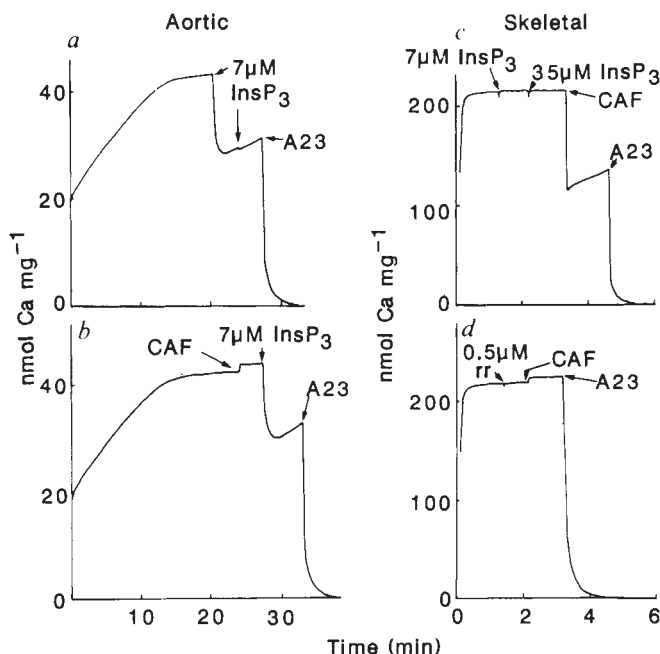


Fig. 1 Comparison of the ability of InsP_3 and caffeine to induce calcium release from SR vesicles made from smooth and striated muscle. Calcium release (downward deflection) was initiated from aortic SR vesicles by addition of $7 \mu\text{M}$ InsP_3 (a), but not by addition of 2.2 mM caffeine (CAF; b). Skeletal muscle SR, by contrast, did not release calcium upon addition of $7 \mu\text{M}$ or $35 \mu\text{M}$ InsP_3 , but did release calcium in the presence of 2.2 mM caffeine (c). The caffeine-induced calcium release from skeletal muscle SR was completely inhibited by $0.5 \mu\text{M}$ ruthenium red (RR) (d). A23 represents the addition of $2 \mu\text{M}$ A23187 (a calcium ionophore used to determine total intravesicular calcium).

Methods. SR vesicles derived from canine aortic muscle and rabbit skeletal muscle were made as described previously^{7,34}. Calcium uptake and release by SR vesicles were monitored by dual wavelength spectrophotometry, using the calcium indicator antipyrilazo III (720, 790 nm). Calcium uptake was initiated by addition of 0.04 vol SR vesicles to reaction media (final concentrations: 100 mM KCl, 0.5 mM MgCl_2 , 1.5 mM Na_2ATP , 5 mM creatine phosphate, 8 U ml^{-1} creatine phosphokinase, 0.4 mM antipyrilazo III, 20 mM HEPES (pH 6.8, 25°C), 5 mM NaN_3 , 0.7 mg ml^{-1} SR). InsP_3 , 2.2 mM caffeine (CAF), $0.5 \mu\text{M}$ ruthenium red (RR), and $2 \mu\text{M}$ A23187 (A23) were added to the reaction media at the times indicated.

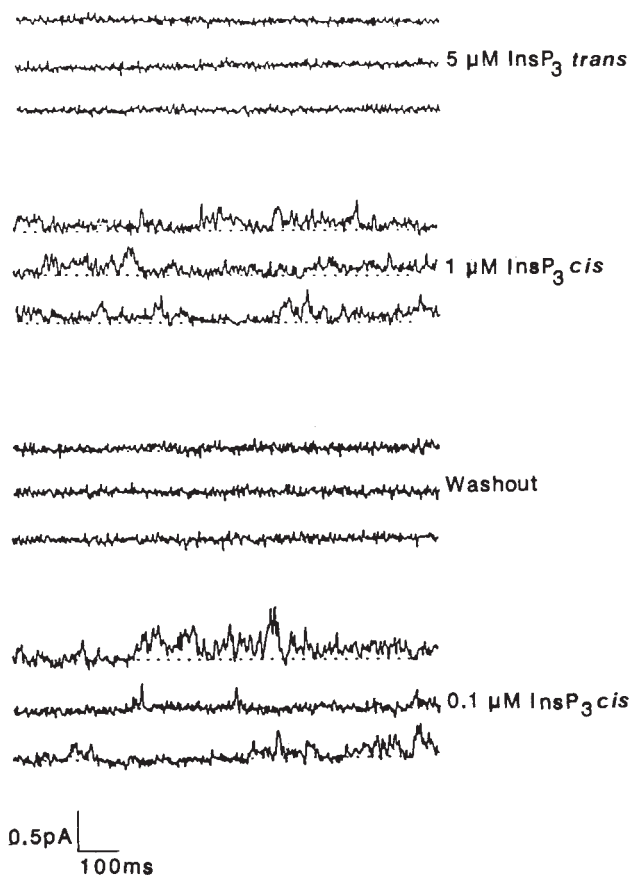


Fig. 2 Inositol 1,4,5-trisphosphate (InsP_3) is effective in opening channels from aortic SR incorporated into planar lipid bilayers only when added to one side of the bilayer. The *cis* side is defined as the side of the bilayer into which the vesicles were added and it has been deduced from these experiments that the *cis* chamber is the cytoplasmic side of the SR. The *trans* side is the lumen of the SR. When $5 \mu\text{M}$ InsP_3 was added to the *trans* side of the bilayer, channel activity was not observed (top three traces). Addition of $1 \mu\text{M}$ InsP_3 to the *cis* side induced channel activity that ceased after the *cis* solution was exchanged with an InsP_3 -free solution (middle six traces). Individual channel fluctuations were too rapid to be resolved in this experiment. Subsequent addition of $0.1 \mu\text{M}$ InsP_3 to the *cis* side of the bilayer showed that a 10-fold lower concentration of InsP_3 was also able to induce channel activity (bottom three traces).

Methods. Bilayers made from phosphatidylethanolamine (75%; Avanti Polar Lipids) and phosphatidylserine (25%; Avanti Polar Lipids) were formed by painting a solution of lipid in decane across a $100\text{-}\mu\text{m}$ hole in a Teflon sheet that bisected a Lucite chamber. Aortic SR vesicles were pre-tested in a vesicle release assay (Fig. 1 and ref. 7) before being used for bilayer experiments. Channel incorporation into the bilayer was followed as described previously²⁶. Briefly, vesicles were added to the *cis* chamber which contained 300 mM N-methyl D-glucamine Cl, 10 mM HEPES, 5 mM CaCl_2 , 0.1 mM EGTA, pH 7.3. The *trans* chamber contained 53 mM CaOH, 250 mM HEPES, pH 7.3 and was held at virtual ground. SR vesicles contain chloride channels that were used as indicators of vesicle fusion with the bilayer²⁶. After the appearance of a single chloride channel in the bilayer the *cis* chamber was perfused with eight volumes of a chloride-free solution (250 mM HEPES-Tris, 1 mM EGTA, 0.5 mM CaCl_2 , pH 7.3) leaving the calcium on the *trans* side of the bilayer as the only small ion present in the system. Bilayer currents were monitored at a fixed holding potential (0 mV) and channel openings indicating divalent cation movement from the *trans* to the *cis* chamber are shown as upward deflections from zero current, which is indicated by a dotted line. Data were stored on a digital tape recorder. For computer analysis data were filtered to 300 Hz with an eight-pole Bessel filter and digitized at 2 kHz .

mM) did not initiate calcium release from these vesicles (refs 7, 20, and data not shown). In addition, the presence of inositol 1,3,4,5-tetrakisphosphate did not enhance the extent of calcium release induced by InsP_3 ²⁰. Consistent with a previous report²¹, the InsP_3 analogue inositol 2,4,5-trisphosphate could initiate calcium release from aortic SR vesicles, though with a much lower sensitivity ($K_{0.5} = 1 \mu\text{M}$ for inositol 1,4,5-trisphosphate; $K_{0.5} = 20 \mu\text{M}$ for inositol 2,4,5-trisphosphate). In the light of

previous reports that showed caffeine-induced calcium release from permeabilized rat smooth muscle cells^{22,23}, the inability of caffeine ($1\text{--}50 \text{ mM}$) to induce calcium release from the aortic SR vesicles could be explained if the caffeine-sensitive calcium channel is present in another region of the SR not isolated in this study or if the caffeine-sensitive component of the channel²⁴ is altered/lost during vesicle isolation.

When aortic SR vesicles were incorporated into planar lipid

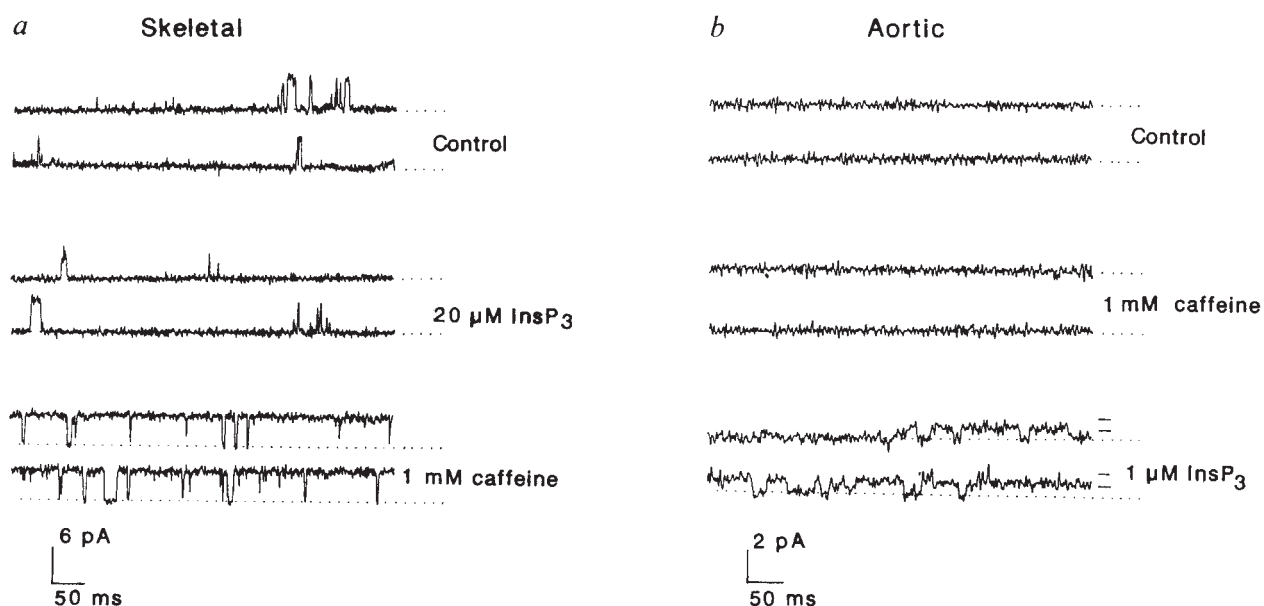


Fig. 3 Comparison of SR vesicles from skeletal and aortic muscle incorporated into planar lipid bilayers. Zero current is indicated by dotted lines and channel openings are upward deflection from zero current. *a*, Skeletal SR vesicle channels were observed at a free Ca^{2+} concentration of 100 nM with a probability of finding the channel open (P_0) of <2% (top two traces). Addition of 20 μM InsP_3 had no effect on P_0 (middle two traces) whereas subsequent addition of 1 mM caffeine increased P_0 to 87% (bottom two traces). *b*, Aortic SR vesicle channels were only observed in the presence of InsP_3 . In 100 nM free Ca^{2+} (top two traces) or in the presence of 1 mM caffeine (middle two traces) bilayer currents were stable and appeared as if no channels were present. Subsequent addition of 1 μM InsP_3 increased the P_0 to 37% (bottom two traces). Two active channels are present in this bilayer. Short lines to the right of the traces indicate single current steps. *c*, Comparison of the single-channel current-voltage relationship of the channels from skeletal, cardiac, and aortic SR vesicles. Lines are drawn to connect the points. Slope conductance between -20 and 0 mV is 125 pS for the skeletal SR channel, 105 pS for the cardiac SR channel, and 10 pS for the aortic SR channel. Bilayer formation and solutions as described in Fig. 2. Skeletal and cardiac SR vesicles were prepared as described previously^{34,35}. P_0 was determined from at least 90 s of continuously recorded data. Single-channel currents were determined in at least three experiments for each muscle type. The points in 3c are larger than the standard error of the means. The free Ca^{2+} concentration was calculated using a Ca-EGTA association constant of 1.011×10^7 (ref. 36).

bilayers single-channel currents were observed in the presence of InsP_3 . In the absence of InsP_3 , bilayer currents were stable and appeared as if no channels were present. In addition, InsP_3 was effective only when it was added to one side of the bilayer. The *cis* side of the membrane was the side of the bilayer into which the vesicles were added. The vesicles inserted into the bilayer so that the cytoplasmic surface faced the *cis* chamber and the luminal surface of the SR faced the *trans* chamber. When 5 μM InsP_3 was added to the *trans* side of the bilayer (Fig. 2, top three traces) channel activity was not observed. Addition of 1 μM InsP_3 to the *cis* side activated a channel (Fig. 2, traces 4-6) and the activity ceased when InsP_3 was washed out of the *cis* chamber (Fig. 2, traces 7-9). Subsequent addition of 0.1 μM InsP_3 again opened channels (Fig. 1, bottom three traces). Because InsP_3 -induced activation was seen after extensive perfusion of the *cis* side of the bilayer, any cofactors necessary for activation would have to be tightly associated with the bilayer²⁵.

The properties of single channels incorporated into planar lipid bilayers had properties predicted from vesicle release experiments. The bilayer experiments determined the response of individual channels, whereas the vesicle release experiments

averaged the response over a large number of channels. In the bilayer the channel from aortic SR and the channel from skeletal muscle SR are activated by different compounds (Fig. 3) as previously shown in vesicle release experiments (Fig. 1). With rabbit skeletal SR channels, up to 20 μM InsP_3 had no effect on basal activity (P_0 , the probability of a channel being open, remained at <2%; Fig. 3a, middle two traces). Other investigators have shown that up to 125 μM InsP_3 had no effect on striated muscle SR channel activity^{26,27}. Addition of 1 mM caffeine increased the P_0 of the channels from skeletal SR from 1% to 87% (Fig. 3a, bottom two traces). Similar results were obtained with channels from canine cardiac muscle SR (data not shown). Aortic SR channels were not activated by 1 mM caffeine (Fig. 3b). However, subsequent addition of 1 μM InsP_3 activated a channel from the aortic SR preparation (Fig. 3b, bottom two traces).

Adenine nucleotides also activate calcium channels in skeletal SR²⁶⁻²⁹ and increase the InsP_3 -induced calcium release from permeabilized aortic cells³. We found that with 100 μM AMP-PCP, an analogue of ATP that is poorly hydrolysable, the P_0 of the InsP_3 -gated channels from aortic SR increased 2.0 ± 0.3 fold ($\bar{x} \pm \text{s.e.m.}$, $n = 4$) in the presence of InsP_3 but did not open

Table 1 The effects of inhibitors on calcium release from SR vesicles and on InsP_3 -gated channel activity in the bilayer

Addition	Aortic SR		Skeletal SR	
	Vesicles*	Bilayer†	Vesicles†	Bilayer‡
None	24	100	38	100
Ruthenium Red				
1 μM	22	100	0	0
20 μM	19	—	0	—
Heparin				
10 $\mu\text{g ml}^{-1}$	8	21	37	—
1 mg ml^{-1}	0	—	33	—
Con A				
100 $\mu\text{g ml}^{-1}$	22	—	36	—
1 mg ml^{-1}	23	—	37	—

* Values are percentage release. Aortic SR vesicles were actively loaded with calcium for 7 min (as in Fig. 1 and ref. 7 with addition of the specified compound during the pre-incubation period), then 7 μM InsP_3 was added, followed 1.5 min later by 4 μM A23187. The data represent the percentage of the total intravesicular calcium (as determined by addition of A23187) that could be released by 7 μM InsP_3 . Calcium content at the time of InsP_3 addition was $40 \pm 1 \text{ nmol mg}^{-1}$ under all conditions. † Values are percentage release. The same assay condition were used as for InsP_3 -induced calcium release from aortic SR, except that skeletal SR vesicles⁷ were used and calcium release was initiated by addition of 4.6 mM caffeine. The data represent percent of total intravesicular calcium that could be released by 4.6 mM caffeine. The calcium content at the time of caffeine addition was $125 \pm 2 \text{ nmol mg}^{-1}$ under all conditions. ‡ Values are channel activity relative to 'control' which is defined as addition of InsP_3 only for aortic SR or caffeine for skeletal SR. To compare different experiments it is necessary to arbitrarily assign the 'control' level of activity to 100%. Results from at least three experiments were averaged. Channels were incorporated into bilayers and analysed as described in Fig. 2.

channels by itself. In one experiment, 100 μM ATP activated the InsP_3 -gated channels to the same extent as AMP-PCP in the presence of InsP_3 , but did not open channels by itself. This contrasts with striated muscle SR calcium channels where either ATP or AMP-PCP alone can open the channel^{26,27}.

The single-channel currents also differ among the muscle types (Fig. 3c). Using the same experimental conditions, the single-channel conductance of the InsP_3 -gated channel from aortic SR was $\sim 10 \text{ pS}$ (slope conductance calculated between 0 and -20 mV ; Fig. 3c, circles). This is approximately ten times lower than the conductance calculated for the caffeine-sensitive calcium channels from cardiac SR (105 pS; Fig. 3c, triangles) and skeletal SR (125 pS; Fig. 3c, squares). Our results are comparable to previous estimates of single-channel conductance for InsP_3 -gated channels in the outer membrane of T lymphocytes¹⁰ and for calcium-gated calcium channels from skeletal SR²⁶ and cardiac SR²⁷. Over the limited voltage range studied (0 to -40 mV), the InsP_3 -gated channels do not appear to be voltage-dependent.

The effects of channel inhibitors reveals additional differences between aortic and skeletal SR calcium channels. The hexavalent cation ruthenium red is known to inhibit calcium release from SR vesicles derived from striated muscle and to inhibit calcium channels from striated muscle SR in bilayers (Table 1 and refs 27, 28). We found that 1–20 μM ruthenium red failed to inhibit either the InsP_3 -induced calcium release from aortic SR or the activity of InsP_3 -gated channels from aortic SR in the bilayer (Table 1). Ruthenium red seems to inhibit a large conductance (that is, $\sim 100 \text{ pS}$) channel regardless of the compound used to open the channel^{26,30}, but it does not block the low conductance channel from aortic SR.

Heparin has been shown to bind to an InsP_3 -receptor found in cerebellum³¹ and to inhibit tension development in saponin-permeabilized smooth muscle³² but was ineffective in altering

calcium release in intact skeletal muscle fibres³³. Therefore, we examined the effect of heparin on InsP_3 -gated calcium release and on the activity of channels incorporated into bilayers. Heparin (10 $\mu\text{g ml}^{-1}$) markedly reduced InsP_3 -induced calcium release from aortic SR, whereas concentrations of up to 1 mg ml^{-1} had no effect on caffeine-induced calcium release from skeletal SR vesicles (Table 1). These concentrations of heparin had no effect on aortic or skeletal SR calcium accumulation (Table 1). Inhibition by heparin was also observed in the aortic SR channels incorporated into planar lipid bilayers when heparin was added to the *cis* side of the membrane (Table 1). The InsP_3 receptor of cerebellum also binds concanavalin A³¹ (con A); but concentrations of con A as high as 1 mg ml^{-1} failed to alter either InsP_3 -induced calcium release from aortic SR or the caffeine-induced calcium release from skeletal SR (Table 1). The similarity of the concentration of heparin ($\sim 10 \mu\text{g ml}^{-1}$) needed to inhibit InsP_3 binding to a protein in cerebellum³¹, InsP_3 -induced tension development in smooth muscle³², InsP_3 -induced calcium release from aortic SR vesicles, and InsP_3 -gated channel activity in the bilayer suggests that InsP_3 and heparin are binding to the same receptor.

In summary, aortic SR vesicles were shown by flux measurements and single-channel current measurements to contain an InsP_3 -gated calcium permeation pathway. The single-channel analyses provide direct evidence that InsP_3 activates a channel in aortic SR vesicles at physiologically relevant concentrations of InsP_3 ¹⁹. The properties of the aortic SR channel seem to be quite different from the calcium-gated calcium release channels of skeletal and cardiac SR. The ability to open channels at very low InsP_3 concentrations and to block the channels with concentrations of heparin known to inhibit tension development in permeabilized smooth muscle³² provide strong support for the hypothesis that InsP_3 has a major role in coupling receptor activation and contraction in smooth muscle.

We thank E. M. Adler, J. E. Brown, L. B. Cohen, A. M. Katz, D. H. Kim, I. B. Levitan, C. Miller and R. Sloboda for comments on the manuscript. This work was supported by the NIH. BEE is a PEW Scholar in the Biomedical Sciences.

Received 6 October; accepted 27 October 1988.

- Abdel Latif, A. *Pharmac. Rev.* **38**, 227–272 (1986).
- Berridge, M. J. *exp. Biol.* **124**, 332–335 (1986).
- Smith, T., Smith, L. & Higgins, B. J. *biol. Chem.* **260**, 14413–14416 (1985).
- Joseph, S. & Williamson, J. J. *biol. Chem.* **261**, 14658–14664 (1984).
- Muallem, S., Schoefield, M., Pandolf, S. & Sachs, G. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4433–4437 (1985).
- Adunyah, S. & Dean, W. J. *biol. Chem.* **261**, 13071–13075 (1985).
- Watrass, J. & Benevolensky, D. *Biochim. biophys. Acta* **931**, 354–363 (1987).
- Suematsu, E., Hirata, M., Hashimoto, T. & Kuriyama, H. *Biochem. biophys. Res. Commun.* **120**, 481–485 (1984).
- O'Rourke, F., Halenda, S., Zavoica, G. & Feinstein, M. J. *biol. Chem.* **260**, 956–962 (1985).
- Kuno, M., Goronzy, J., Weyand, C. & Gardner, P. *Nature* **323**, 269–273 (1986).
- Kuno, M. & Gardner, P. *Nature* **326**, 301–304 (1987).
- Benevolensky, D., Menshikova, E., Watras, J., Levitsky, D. & Ritov, V. *Biomed. biochim. Acta* **8/9**, S393–S398 (1987).
- Volpe, P., Salvati, G., Di Virgilio, V. & Pozzan, T. *Nature* **316**, 347–349 (1985).
- Vergara, J., Tsien, R. & Delay, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6352–6356 (1985).
- Nosek, T., Williams, M., Zeigler, S. & Godt, R. *Am. J. Physiol.* **250**, C807–C811 (1986).
- Movesian, M., Thomas, A. & Williamson, J. *FEBS Lett.* **185**, 328–332 (1985).
- Scherer, N. & Ferguson, J. *Biochem. biophys. Res. Commun.* **128**, 1064–1070 (1985).
- Ritov, V., Menshikova, E. & Khodorov, B. *Biologicheskii Membran* **3**, 1122–1129 (1986).
- Walker, J. W., Somlyo, A. V., Goldman, Y. E., Somlyo, A. P. & Trentham, D. R. *Nature* **327**, 249–252 (1987).
- Watrass, J., Benevolensky, D. & Childs, C. J. *molec. cell. Cardiol.* (in the press).
- Irvine, R. F., Letcher, A. J., Lander, D. J. & Berridge, M. J. *Biochem. J.* **240**, 301–304 (1986).
- Yamamoto, H. & van Breemen, C. *Biochem. biophys. Res. Commun.* **130**, 270–274 (1985).
- Kobayashi, S., Kanaide, H. & Nakamura, M. J. *biol. Chem.* **261**, 15709–15713 (1986).
- Rubtsov, A. M. & Murphy, A. J. *Biochem. biophys. Res. Commun.* **154**, 462–468 (1988).
- Ewald, D. A., Williams, A. & Levitan, I. B. *Nature* **315**, 503–506 (1985).
- Smith, J., Coronado, R. & Meissner, G. J. *gen. Phys.* **88**, 573–588 (1986).
- Rousseau, E., Smith, J., Henderson, J. & Meissner, G. *Biophys. J.* **50**, 1009–1014 (1986).
- Smith, J., Coronado, R., Meissner, G. *Nature* **316**, 446–449 (1985).
- Meissner, G., Darling, E. & Eveleth, J. *Biochem. J.* **25**, 236–244 (1986).
- Suarez-Isla, B. A. *et al. Biophys. J.* **53**, 467a (1988).
- Supattapone, S., Worley, P., Baraban, J. & Snyder, S. J. *biol. Chem.* **263**, 1530–1534 (1988).
- Kobayashi, S., Somlyo, A. V. & Somlyo, A. P. *Biochem. biophys. Res. Commun.* **153**, 625–631 (1988).
- Pape, P. C., Konishi, M., Baylor, S. M. & Somlyo, A. P. *FEBS Lett.* **235**, 57–62 (1988).
- Kim, D. H., Ohnishi, S. & Ikemoto, N. J. *biol. Chem.* **258**, 9662–9668 (1983).
- Kranias, E., Schwartz, A. & Jungmann, R. *Biochim. biophys. Acta* **709**, 28–37 (1982).
- Fabiato, A. & Fabiato, F. *J. Physiol. (Paris)* **75**, 463–505 (1979).

Opening of dihydropyridine calcium channels in skeletal muscle membranes by inositol trisphosphate

Janeen Vilven & Roberto Coronado*

Department of Physiology and Molecular Biophysics,
Baylor College of Medicine, Houston, Texas 77030, USA

In many non-muscle cells^{1,2}, D-inositol 1,4,5-trisphosphate (InsP₃) has been shown to release Ca²⁺ from intracellular stores, presumably from the endoplasmic reticulum. It is thought to be a ubiquitous second messenger that is produced in, and released from, the plasma membrane in response to extracellular receptor stimulation. By analogy, InsP₃ in muscle cells has been postulated to open calcium channels in the sarcoplasmic reticulum (SR) membrane, which is the intracellular Ca²⁺ store that releases Ca²⁺ during muscle contraction^{3,4}. We report here that InsP₃ may have a second site of action. We show that InsP₃ opens dihydropyridine-sensitive Ca²⁺ channels in a vesicular preparation of rabbit skeletal muscle transverse tubules⁵⁻¹⁰. InsP₃-activated channels and channels activated by a dihydropyridine agonist in the same preparation have similar slope conductance and extrapolated reversal potential

and are blocked by a dihydropyridine antagonist. This suggests that in skeletal muscle, InsP₃ can modulate Ca²⁺ channels of transverse tubules from plasma membrane, in contrast to the previous suggestion that the functional locus of InsP₃ is exclusively in the sarcoplasmic reticulum membrane.

It has been suggested that InsP₃ serves as the 'chemical messenger' of excitation-contraction coupling³⁻⁴. Produced in the tubular membrane in response to an action potential, InsP₃ would then diffuse to the SR to activate Ca²⁺ release by opening Ca²⁺ channels. This hypothesis has been met with a large number of conflicting reports from studies in both skinned fibres and vesicular preparations¹¹⁻¹⁸. Two reports suggested that, in addition to the SR, the state of the transverse tubular membrane is important to the effects of InsP₃ in skeletal muscle. Skinned fibres with 'depolarized' t-tubules have been shown to be more sensitive to InsP₃ than fibres with 'polarized' tubules¹⁶. The InsP₃ response may therefore directly involve depolarization of tubules which become sealed and spontaneously 'polarized' during the skinning procedure¹⁷. These observations imply that, more than the SR itself, it is the junction between the transverse tubules and the SR membranes that is the site of the InsP₃ response. We therefore decided to investigate the possible electrophysiological effects of phosphoinositides on the dihydropyridine-sensitive Ca²⁺ channel of the transverse tubular membrane of rabbit skeletal muscle⁵⁻¹⁰. Figure 1 shows four separate recordings of Ca²⁺ channels activated by 10 µM InsP₃, or 1 µM Bay K8644 dihydropyridine agonist, or both, at a

* To whom correspondence should be addressed.

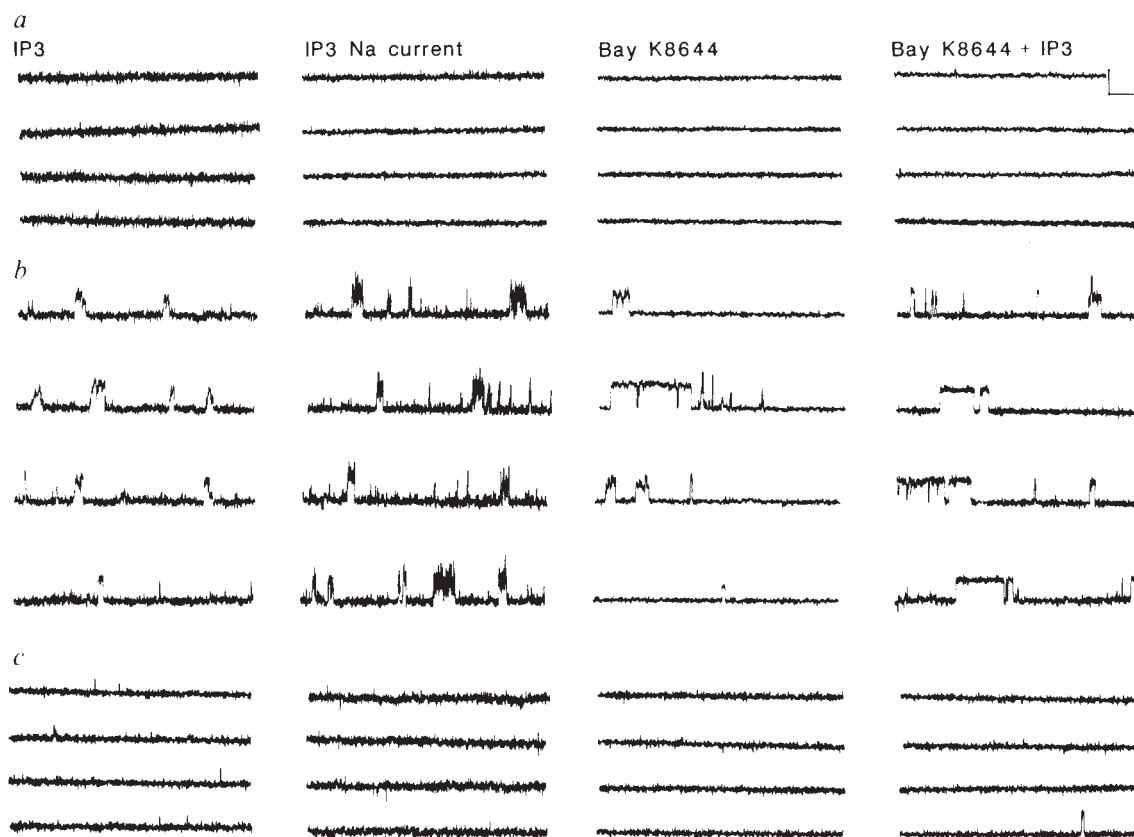


Fig. 1 T-tubule Ca²⁺ channels activated by InsP₃ and Bay K8644. Four separate experiments are divided into segments: *a*, in the presence of *cis* t-tubule protein without InsP₃ or Bay K8644, *b*, following *cis*-internal addition of 10 µM InsP₃ or 1 µM Bay K8644 and appearance of channels and, *c*, following *cis*- and *trans*-external addition of 20 µM nitrendipine.

Methods. Current carrier was *cis* 100 mM BaCl₂, 0.05 M NaCl, *trans* 0.05 M NaCl in all cases except that the InsP₃ Na⁺ current contained *cis* 250 mM NaCl, *trans* 0.05 M NaCl, 2 µM free Ca²⁺. Holding potential was +20 mV. Buffering solutions are given in Fig. 2. Records were filtered at 0.05 kHz corner frequency on an eight-pole Bessel, digitized at 1 point ms⁻¹, and analysed by computer. Lipid bilayers were cast from an equimolar mixture of phosphatidylethanolamine and phosphatidylserine dissolved in decane as described⁸. T-tubule vesicles were prepared from rabbit back and leg white muscle as described⁸. Light muscle microsomes sedimenting at 10/20% sucrose interface were used in all experiments. [³H]PN200-110 binding was routinely 30–40 pmol per mg protein. Phospholipids were from Avanti Polar Lipids Alabama, Bay K8644 was a gift from Miles Pharmaceutical Div. Connecticut, InsP₃ was purchased from Sigma. For single-channel recording reagents were analytical grade (Johnson Matthey).

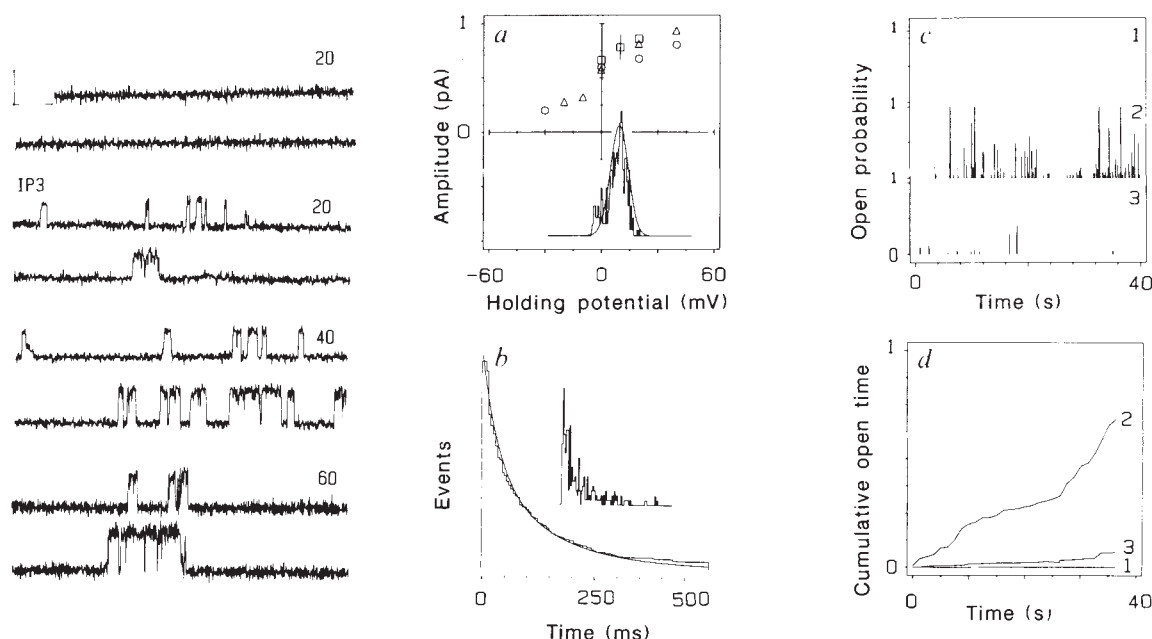


Fig. 2 Single-channel characteristics of InsP₃ activated channels. (Left) Barium current through Ca²⁺ channels activated by 10 μM InsP₃ at +20 mV, +40 mV and +60 mV. Control records are in the absence of InsP₃ after *cis* addition of 3 μg of t-tubule vesicular protein. **a**, Current-voltage curve for single channels activated by 10 μM InsP₃ (circles); 1 μM Bay K8644 plus 10 μM InsP₃ (triangles); and 1 μM Bay K8644 (squares). Internal solution was 0.1 M BaCl₂, 0.05 M NaCl, 0.01 M HEPES-Tris. External solution was 0.05 M NaCl, 0.01 M HEPES-Tris; pH 7.0. Bars correspond to 2 s.d. of the data set with the largest error. Inset correspond to 244 events at +20 mV fitted to a single gaussian curve with mean 0.7 pA, s.d. = 0.45 pA. **b**, Cumulative time histogram of the same 244 events above plotted as % (events longer than time, *t*) against *t*. Smooth lines corresponds to the sum of two exponentials with time constants 53 ms and 117 ms; 51% of total events are fit by the distribution of brief events. Inset corresponds to the time distribution of each of the events sorted at 5-ms intervals (500 ms is the total width of the inset histogram). **c**, **d**, Diary of channel activity following additions of InsP₃ and nitrendipine. Trace 1, 40 s of continuous activity in control with t-tubule protein without InsP₃; trace 2, following *cis* addition of 10 μM InsP₃; trace 3, after 20 μM *trans* nitrendipine are shown in (c) on a sweep per sweep basis, or in (d) as a cumulative sum. The probability of finding one or more open channels on a segment of 500 ms is plotted on the y axis. Consecutive segments in time are given in c as bars of value 0 to 1. Empty segments are periods with no openings. In d, probabilities per segment are added and the cumulative sum plotted as a function of time.

holding potential of +20 mV. Recordings were made in planar bilayers using the vesicle/bilayer fusion technique. Except for when measuring the Na⁺ current (IP₃ Na⁺ current), the current carrier was 100 mM BaCl₂ in the *cis* side (the chamber bathing the internal side of the dihydropyridine Ca²⁺ channel)^{9,10}. Therefore current flow at 0 mV is in the outward direction and this is shown upwards in all cases ($E_{Ba} < -100$ mV, nominally minus infinity; $E_{Na} = 0$ mV, maintained by 50 mM NaCl on both sides; $E_{Cl} = +35$ mV; *trans*-external side held at ground potential). Outward Na⁺ current was recorded in *cis* 0.25 M/*trans* 0.05 M NaCl with 2 μM free Ca²⁺ ($E_{Na} = -35$ mV; $E_{Cl} = +35$ mV). In the absence of InsP₃ or Bay K8644, there were no measurable openings following formation of bilayers and addition of t-tubule protein to the *cis* chamber (a traces in Fig. 1). In our experience, spontaneous openings of Ca²⁺ channels at constant 0 mV and in the absence of dihydropyridine agonist is extremely low. In the vesicular preparation of rabbit transverse tubules the agonist-free activity is equivalent to an average open channel probability, $P_o < 0.001$. In 96% of the membrane preparations used in various studies⁵⁻¹⁰ ($n = 50$) this value was actually zero. InsP₃ (10 μM), 1 μM Bay K8644, or both added at the same time, induced openings of channels carrying Ba²⁺ current with an approximate amplitude of 0.5 pA at 0 mV and a slope conductance of about 10 pS in all cases (b traces in Fig. 1). Activation of channels by InsP₃ without Bay K8644 has been replicated 24 times in six separate preparations of t-tubules. The average time for onset of channel activity was 15 min with lower and upper limits of 90 s and 40 min, respectively. InsP₃ also activated channels carrying Na⁺ current (IP₃ Na current) which, under the conditions tested, had a mean current of 0.6 pA at 0 mV and a slope conductance of approximately 20 pS at the

same potential. This value agrees well with the conductance of 24 pS that we reported for the Na⁺ current of Bay K8644-activated Ca²⁺ channels in the same preparation⁶⁻⁷. The traces labelled c in Fig. 1 show that all channel activities initiated by InsP₃ or Bay K8644 are blocked by 20 μM nitrendipine, a dihydropyridine antagonist. This result confirmed that InsP₃-activated channels are indeed pharmacologically related to dihydropyridine Ca²⁺ channels in this preparation.

We further examined the single-channel properties of InsP₃ activated channels (Fig. 2). In mixed Ba²⁺ and Na⁺ chloride salt solutions, unitary amplitudes increased at positive potentials and decreased at negative potentials with reversal potential closer to E_{Ba} and more than 40 mV away from E_{Na} or 80 mV away from E_{Cl} (Fig. 2a). Slope conductance calculated from the best fit line was 11 pS (s.d. = 2.1 pS) for activation by InsP₃ (circles), 10 pS (s.d. = 1.5 pS) for activation by Bay K8644 (squares), and 12 pS (s.d. = 1.7 pS) for activation by Bay K8644 and InsP₃ added together (triangles). In each case the extrapolated reversal potential was more negative than -40 mV corresponding to a Goldman permeability ratio $P_{Ba}/P_{Na} > 20$. Chloride permeability was found to be negligible in each case. The inset in Fig. 2a corresponds to the amplitude distribution of 244 events at +20 mV. The distribution of open times for the same data is shown in Fig. 2b as a percentage of durations longer than time, *t*, plotted as a function of time, *t*, that is, as a cumulative histogram with the first bin at *t* = 0 containing the total 244 events. The inset in Fig. 2b gives the duration of each event. The distribution could be adequately fitted as the sum of two populations of approximately equal size and mean lifetimes 32 and 117 ms. Openings induced by 1 μM Bay K8644 under the same conditions could also be fitted by two exponentials

but with significantly larger means, 53 and 206 ms ($n = 206$ events). A quantitative expression of channel activation by InsP_3 and blockade by nitrendipine is given in Fig. 2c, d. The probability of finding one or more open channels as a function of time was calculated every 250 ms for a total period of 40 s before addition of InsP_3 (trace 1), after channel activation by $10 \mu\text{M}$ InsP_3 (trace 2) and after blockade by $20 \mu\text{M}$ nitrendipine (trace 3). In c, the open probability during each consecutive period of 250 ms is plotted as a bar of magnitude 0 to 1. In Fig. 2d, the same experiment is plotted as the cumulative sum of probabilities in each period. A straight line in Fig. 2d would indicate a level of channel activation in which the nP_o product is invariant with time where n is the number of channels and P_o is the open probability. The trace 2 in Fig. 2c, d indicates that this was roughly so. In eight cases analysed InsP_3 induced a stationary open probability that remained constant for up to 240 s, the longest monitoring period. Hence InsP_3 channels do not inactivate with time. The average probability of one or more open channels (per 250 ms period) was $P_o = 0.014$ ($n = 8$) for $10 \mu\text{M}$ InsP_3 , $P_o = 0.035$ for $1 \mu\text{M}$ Bay K8644 ($n = 4$) and $P_o = 0.145$ ($n = 2$) for both added together which indicates a possible synergism of the two activators. The level of activity in trace 2 compared to that of trace 3 indicates that $20 \mu\text{M}$ nitrendipine blocked InsP_3 channels by approximately 10-fold.

Three separate results suggest that in the t-tubule preparation of skeletal muscle, the Ca^{2+} channel activated by Bay K8644 is the same as that independently activated by InsP_3 . (1) Both channels have a unit conductance of 10 pS in 100 mM Ba^{2+} and a selectivity $P_{\text{Ba}}/P_{\text{Na}} > 20$; (2) $20 \mu\text{M}$ nitrendipine blocks InsP_3 -activated and Bay K8644-activated channels; (3) in the absence of divalent ions in solution, the InsP_3 channel is permeable to Na^+ ions. We have made the same observation in dihydropyridine Ca^{2+} channels^{6,7}. In control experiments $10 \mu\text{M}$ inositol, inositol 1-phosphate, or inositol 1,2-cyclic phosphate, or incubation of vesicles for one hour with 1 mM MgATP failed to activate Ca^{2+} channels. However, the waiting time until channels were activated by InsP_3 was considerably reduced after incubation with MgATP. Also inositol 1,4-bisphosphate could replace InsP_3 in MgATP-treated vesicles, which probably reflects an effect of inositol 1,4-bisphosphate at some intermediary step of the phosphoinositide cascade. Inositol, inositol 1-phosphate or inositol 1,2-cyclic phosphate were ineffective in the presence or absence of MgATP. The fact that InsP_3 had an agonist effect at an 'infinite' dilution of cellular components, and in extremely simplified ionic solutions, is good evidence for a direct effect of InsP_3 on the Ca^{2+} channel. As the channel is inserted in a synthetic membrane that has an area roughly five orders of magnitude larger than a single t-tubule vesicle, all components present in the vesicle that fuses into the planar bilayer, would be immensely diluted into a vast excess of exogenous lipid. It is therefore very unlikely that if InsP_3 acted indirectly through an intermediary receptor, this receptor would in turn be able to locate the channel in the plane of the membrane. The large volume of the chamber in contact with the channel (3 ml on either side) helps eliminate the possibility that InsP_3 may be generating a diffusible activator.

Recent evidence indicates that the dihydropyridine receptor of skeletal muscle may have a dual function: as a Ca^{2+} channel^{19,20}, and as a voltage sensor of t-tubule depolarization responsible for triggering SR Ca^{2+} release²¹. It is also likely that InsP_3 may have several loci of action in excitation-contraction coupling, including that of putative second messenger¹⁸. All our evidence for InsP_3 activation of the Ca^{2+} release channel of rabbit skeletal muscle SR is negative however, and so we suggest that the site of action of InsP_3 in skeletal muscle is either the tubular membrane or the t-tubule/SR junction but not the SR membrane directly. This conclusion is directly supported by the work of Hannon and colleagues¹⁷, which first suggested that InsP_3 could open channels present in the transverse tubular membrane. Finally, the fact that InsP_3 opens Ca^{2+} channels at

0 mV is strong evidence in favour of a shift from inactivated-to-rest states or inactivated-to-open states. Otherwise, we would have not been able to record openings at depolarized potentials as all Ca^{2+} channels are inactivated at the time of membrane fractionation and purification. We suggest that InsP_3 could primarily facilitate the switching of channels away from the inactive state thus serving as an 'endogenous modulator'. Our data show that InsP_3 can mimic the dihydropyridine agonist Bay K8644 by directly opening Ca^{2+} channels although we do not know whether *in vivo* conditions favour this direct agonist effect.

This work was supported by NIH, Grants in Aid from the American Heart Association and the Muscular Dystrophy Association of America and by an Established Investigatorship from the American Heart Association to R.C.

Received 13 May; accepted 31 October 1988

- Berridge, M. J. *exp. Biol.* **124**, 323-335 (1986).
- Berridge, M. A. *Rev. Biochem.* **56**, 159-193 (1987).
- Vergara, J., Tsien, R. Y. & Delay, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6352-6356 (1985).
- Volpe, P., Salvati, G., Di Virgilio, F. & Pozzan, T. *Nature* **316**, 347-349 (1985).
- Affolter, H. & Coronado, R. *Biophys. J.* **48**, 341-347 (1985).
- Coronado, R. & Affolter, H. *J. gen. Physiol.* **87**, 933-953 (1986).
- Coronado, R. & Smith, J. S. *Biophys. J.* **51**, 497-502 (1987).
- Ma, J. & Coronado, R. *Biophys. J.* **53**, 387-395 (1988).
- Valdivia, H. & Coronado, R. *Biophys. J.* **53**, 555a (1988).
- Vilven, J. *et al. Biophys. J.* **53**, 665a (1988).
- Lea, T. J., Griffiths, P. J., Treagear, R. T. & Ashley, C. C. *FEBS Lett.* **207**, 153-161 (1986).
- Nosek, T., Williams, M., Zeigler, S. & Godt, R. *Am. J. Physiol.* **19**, C807-C811 (1986).
- Walker, J. W., Somlyo, A. V., Goldman, Y. E., Somlyo, A. P. & Trentham, D. R. *Nature* **327**, 249-252 (1987).
- Rojas, E., Nassar-Gentina, V., Luxoro, M., Pollard, M. E. & Carrasco, M. A. *Can. J. Physiol. Pharmacol.* **65**, 672-680 (1987).
- Mikos, G. J. & Snow, T. R. *FEBS Lett.* **927**, 256-260 (1987).
- Donaldson, S. K., Goldberg, N. D., Walseth, T. F. & Huettman, D. A. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5749-5753 (1988).
- Hannon, J. D., Lee, N. K. M. & Blinks, J. R. *Biophys. J.* **53**, 607a (1988).
- Suarez-Isla, B. A. *et al. Biophys. J.* **54**, 737-740 (1988).
- Flockerzi, V. *et al. Nature* **323**, 66-68 (1986).
- Smith, J. S. *et al. Biochem.* **26**, 7182-7188 (1987).
- Rios, E. & Brum, G. *Nature* **325**, 717-720 (1987).

A voltage-dependent chloride channel in the photosynthetic membrane of a higher plant

G. Schönknecht*, R. Hedrich, W. Junge* & K. Raschke

Pflanzenphysiologisches Institut, Universität Göttingen, D-3400 Göttingen, FRG

Photophosphorylation in photosynthetic membranes of plants (thylakoid membranes) is driven by an electrochemical potential difference of the proton¹. The light-driven net uptake of protons into the thylakoid lumen is electrically balanced by the motion of other ions^{2,3}. So far no ion channel has been identified in thylakoids other than the proton channel of the ATP synthase⁴ although charge balance would call for one. Patch-clamp studies⁵ on isolated membrane patches from osmotically inflated thylakoids of *Peperomia metallica* revealed a voltage-dependent and anion-selective channel. At 30 mM $[\text{Cl}^-]$ the single-channel conductance was 65 pS, showing ohmic behaviour between -80 and +80 mV. The opening probability was maximal at about +40 mV (inside the thylakoid). Application of voltage steps caused additional superimposed transient channel openings. This chloride channel could provide a mechanism involved in charge-balancing during light-driven proton uptake by thylakoids.

As in previous experiments with impaled microelectrodes^{6,7} giant chloroplasts of *P. metallica* were used. Proto-plasts

* Permanent address: Biophysik, FB Biologie/Chemie, Universität Osnabrück, D-4500 Osnabrück, FRG.

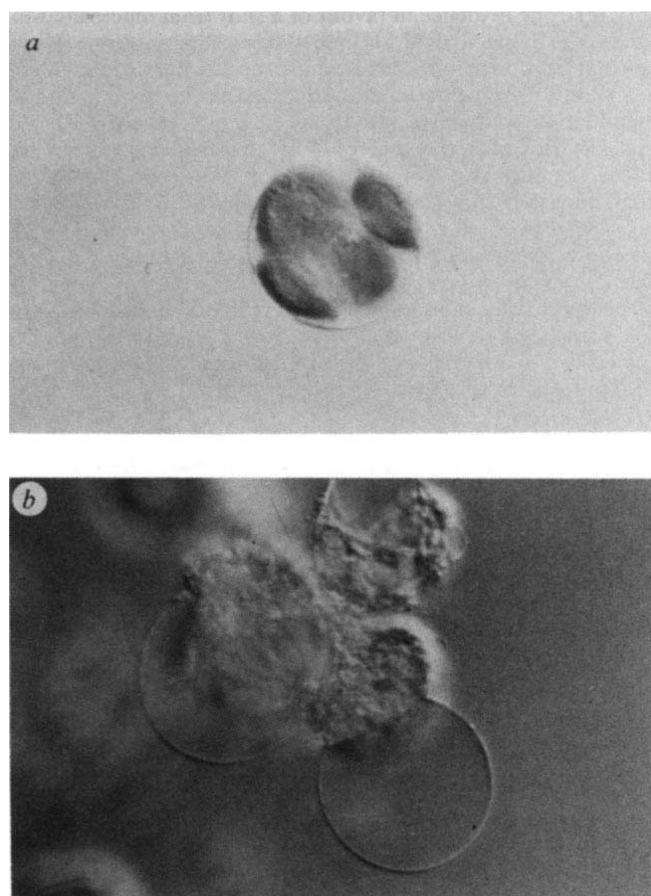


Fig. 1 Light microscopy (Nomarski interference contrast) of *a*, a protoplast (30 μm in diameter) and *b*, osmotically swollen thylakoids ('blebs'; same scale) of *Peperomia metallica*.

Methods. Leaf slices were incubated in 2% cellulase Onozuka R-10 and 1% Mazerocyme R-10 (Yakult Honsha, Tokyo, Japan), 0.5% bovine serum albumin, 0.27 M sorbitol, 1 mM CaCl_2 for about 1 h. Protoplasts were washed with 0.35 M sorbitol, 1 mM CaCl_2 , and separated from debris by filtration through 200- μm and 25- μm nylon nets followed by centrifugation, 7 min at 100g. The protoplasts were suspended in wash medium, stored on ice and used within 4 h. Protoplast suspensions (5–10 μl) were rapidly mixed with a 100-fold volume of bath solution (20 mM KCl, 5 mM MgCl_2 , 2 mM MOPS-KOH, pH 6.9 (MOPS is 3-(N-morpholino) propanesulphonic acid)) in the recording chamber⁹. After 15 min of osmotic swelling in the dark, plasma membranes, vacuoles and thylakoid envelopes had ruptured and thylakoids formed large blebs. The chamber was perfused with bath solution and blebs adhering to the bottom of the chamber were used for patch-clamp studies. All experiments were performed in 20 mM KCl, 5 mM MgCl_2 , 2 mM MOPS-KOH, pH 6.9 inside the pipette plus 1 mM CaCl_2 , unless indicated otherwise. Solutions were changed by bath perfusion.

were isolated from leaves (Fig. 1*a*) and osmotically shocked. This caused the rupture of the plasma membrane, the vacuole and the chloroplast envelope. Within 15 min of incubation several thylakoids were swollen to form large spherical membrane vesicles ('blebs') with up to 40 μm diameter (Fig. 1*b*). The thylakoid origin of the bleb membrane has been established by chlorophyll fluorescence, photochemical redox reactions sensitive to external electric fields⁸, as well as by light-driven⁹ and electric field-driven ATP synthesis¹⁰.

High-resistance seals (>10 G Ω) were formed between the fire-polished tip of patch pipettes and thylakoid membranes. They allowed single-channel recordings either in the thylakoid-attached configuration or, after withdrawal of the pipette, in the inside-out patch configuration⁹ (with the intrathylakoid mem-

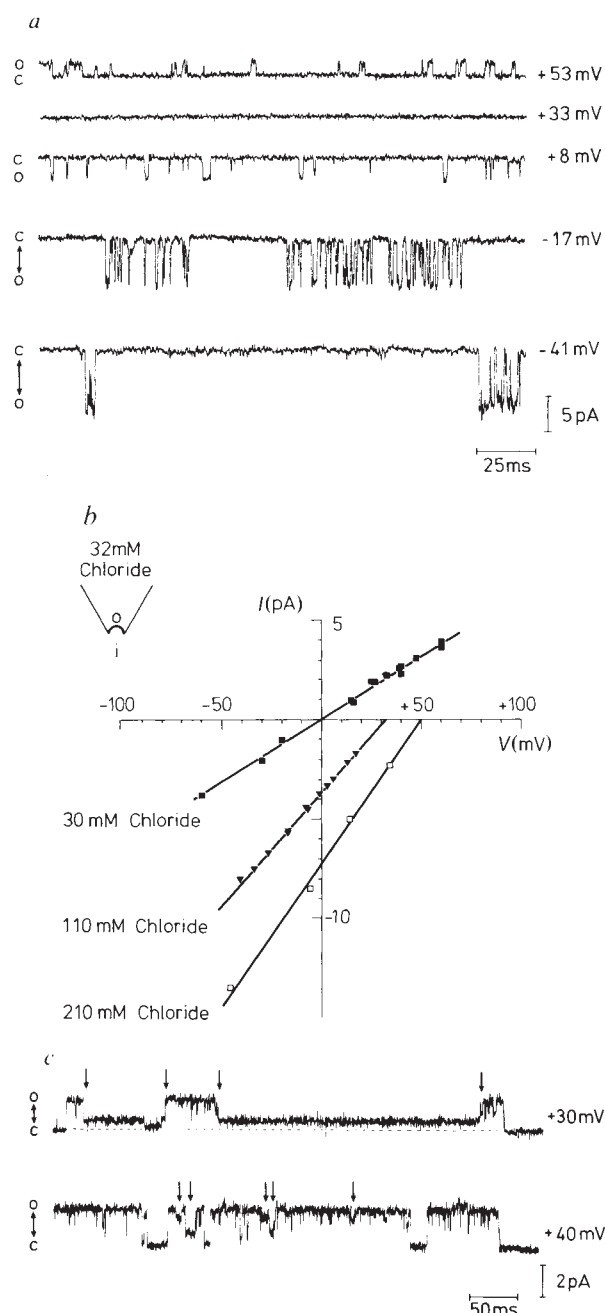
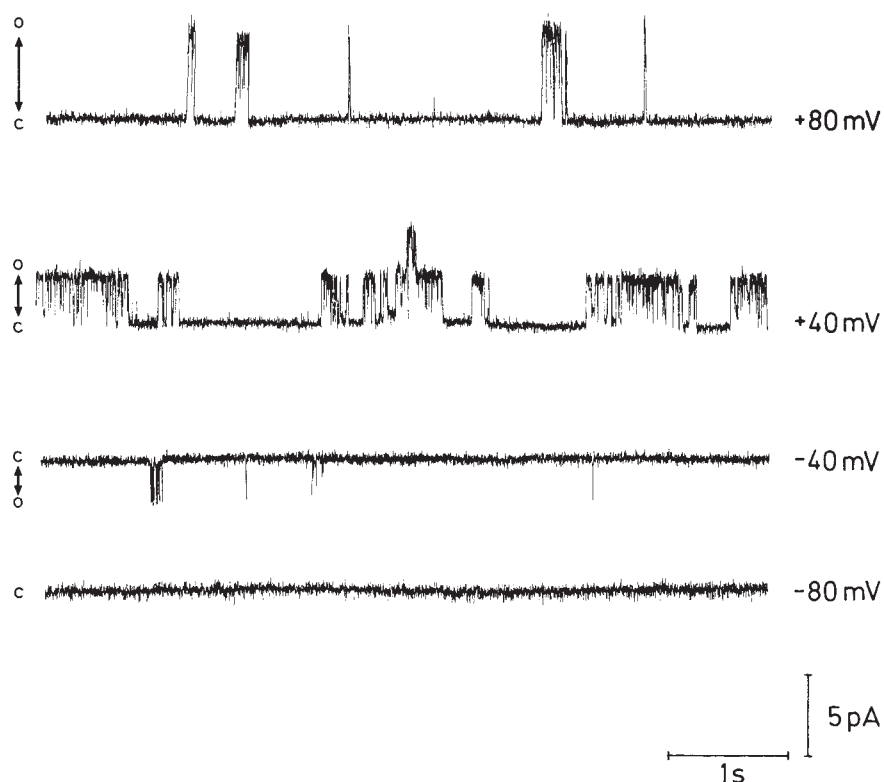


Fig. 2 Selectivity and current-voltage relationship of ion channels in the thylakoid. *a*, Single-channel recordings of chloride channels in inside-out patches (the intrathylakoid membrane side faced the bath solution) under asymmetrical KCl concentrations, 100 mM instead of 20 mM KCl in the bath. Upward current deflections represent Cl^- influx into the thylakoid, whereas downward deflections represent Cl^- efflux. *b*, Current-voltage relationships of the chloride channel at three different KCl gradients. Starting at 20 mM KCl, the concentration in the bath was raised to 100 or 200 mM KCl (background 5 mM MgCl_2 plus 1 mM CaCl_2 inside the pipette). The curves were fitted by linear regression. *c*, Subconductance levels of the chloride channel at holding potentials of +30 mV and +40 mV. All potentials given refer to the intrathylakoid side. **Methods.** Currents were recorded while the membrane potential was clamped to various positive or negative values or while double-voltage steps were applied. The first step, chosen to activate the channel (for example +40 mV; see Fig. 4), was followed by a second step to potentials at which the channel normally was closed. Membrane potentials and currents were measured with an EPC-7 patch-clamp amplifier (List Electronics), current records were stored on video tape, filtered at 2 kHz (8-pole Bessel characteristic) and analysed on a PDP 11/73 minicomputer.

Fig. 3 Voltage dependence of the chloride channel. The four current traces demonstrate representative channel activities, during 6-s time periods, when the potential was held at +80, +40, -40 or -80 mV.



brane face exposed to the bath). With a patch area of about $10 \mu\text{m}^2$ almost all patches contained at least one channel (and up to six, see Fig. 4a). When comparing attached configurations and inside-out patches no differences in single-channel behaviour were observed.

The ion selectivity was determined by the reversal of the single-channel currents in asymmetrical KCl solutions (Fig. 2a). The reversal potential followed the Nernst potential for chloride (Fig. 2b), as expected for an anion channel. Of the anions able to permeate thylakoid membranes¹¹ Cl^- is present in the chloroplast at the highest concentration. Thus the physiological role of this channel is that of a chloride channel.

Saturation of the single-channel conductance was observed by raising the Cl^- concentration on the intrathylakoid side (Fig. 2b): 30 mM, 65 ± 3 pS ($n = 14$); 110 mM, 110 ± 4 pS ($n = 4$); 210 mM, 141 pS and 126 pS. A maximal single-channel conductance of 150 pS and a Michaelis constant, K_m , for chloride of 40 mM were calculated, assuming Michaelis-Menten behaviour of the channel¹². The current-voltage relationship of the open channel was ohmic between +80 and -80 mV in symmetric and asymmetric KCl and KNO_3 solutions. Besides the main conductance state, subconductance levels were found (Fig. 2c), showing the same ion selectivity. Furthermore the thylakoid channel exhibited a complex gating behaviour, as the 'flickers' in Figs 2 and 4 demonstrate.

To determine the relative permeability for NO_3^- with respect to Cl^- , the reversal of the single-channel current was measured by replacing KCl by KNO_3 in the bath. The reversal potential shifted to positive values (data not shown), indicating a slightly higher permeability for NO_3^- than for Cl^- , as was reported for chloride channels in animal tissues¹³. Single-channel currents were about 25% higher in the presence of NO_3^- than in the presence of Cl^- .

Continuous illumination of thylakoids acidifies the lumen by about 3 pH units, we found however that variation of pH (between 7.8 and 5.9) at the intrathylakoid side affected neither the single-channel conductance nor the reversal potential.

Highest channel activity (under voltage-clamp) was observed at about +40 mV (inside the thylakoid). The activity decreased

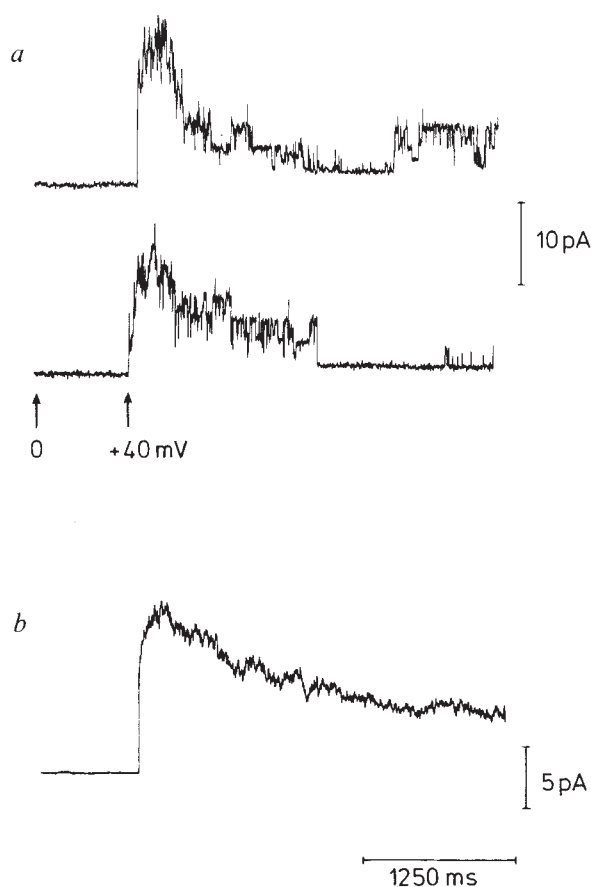


Fig. 4 Single-channel kinetics in response to voltage-step protocols. *a*, From a holding potential of 0 mV the potential was repetitively stepped to +40 mV for 3 s. With the onset of the +40-mV pulse single-channel activity rapidly increased (up to six channels opening simultaneously), followed by a slow decay to a steady state level. *b*, Average of 28 traces as shown in *a*.

as voltages were lowered or raised from +40 mV (Fig. 3). Changes of the holding potential to positive or negative values elicited additional channel openings. Transient channel openings were most pronounced at voltage steps to +40 mV (Fig. 4), a membrane potential comparable to that caused by a single light flash^{3,6}. From the average of 28 traces a half-rise time of less than 100 ms and a half-decay time of about one second were estimated. Generally, channel activity declined within 15–30 min after seal formation.

A voltage-dependent ion channel has been characterized recently in the inner mitochondrial membrane¹⁴. This channel had a similar conductance but was less anion selective than the thylakoid channel.

Because of the complex gating behaviour (sub-levels, flickers) and the decline of channel activity with time, it was not yet possible to study the kinetic properties of the Cl⁻ channel in thylakoids in more detail.

Earlier studies on passive ion movements associated with light-induced proton uptake into thylakoids gave rise to suggestions that electrical compensation was caused by Cl⁻ influx^{2,15}, or Mg²⁺ efflux¹⁶, or both¹⁷. We did not observe K⁺ or Mg²⁺ channels in solutions containing 5 mM MgCl₂ and up to 200 mM KCl. The described chloride channel would allow a rapid,

voltage-dependent uptake of Cl⁻ into thylakoids, and thus an electrical compensation of the light-driven proton uptake.

We thank I. Baumann for technical help, W. Lahr for support with computer software, B. Raufeisen for artwork, and W. Hanke, B. U. Keller and E. Neher for critical comments on the manuscript. Financial support was provided by the Deutsche Forschungsgemeinschaft.

Received 18 August; accepted 4 November 1988.

1. Mitchell, P. *Biol. Rev.* **41**, 445–502 (1966).
2. Rottenberg, H., Grunwald, T. & Avron, M. *Eur. J. Biochem.* **25**, 54–63 (1975).
3. Junge, W. *Curr. Top. Membranes Transp.* **16**, 431–465 (1982).
4. Lill, H., Althoff, G. & Junge, W. *J. Membrane Biol.* **98**, 69–78 (1987).
5. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1980).
6. Vredenberg, W. J. *Proc. 3rd Int. Congr. Photosynthesis* 929–939 (1974).
7. Bulychev, A. A., Andrianov, V. K., Kurella, G. A. & Litvin, F. F. *Nature* **236**, 175–176 (1972).
8. DeGrooth, B. G. & van Gorkom, H. J. *Biochim. biophys. Acta* **35**, 445–456 (1981).
9. Campo, M. L. & Tedeschi, H. *Eur. J. Biochem.* **149**, 511–516 (1985).
10. Gräber, P., Schlödter, E. & Witt, H. T. *Biochim. biophys. Acta* **461**, 426–440 (1977).
11. Schuldiner, S. & Avron, M. *Eur. J. Biochem.* **19**, 227–231 (1971).
12. Läger, P. *Biochim. biophys. Acta* **311**, 423–441 (1973).
13. Bormann, J., Hamill, O. P. & Sakmann, B. *J. Physiol.* **385**, 243–286 (1987).
14. Sorgato, M. C., Keller, B. U. & Stühmer, W. *Nature* **330**, 498–500 (1987).
15. Vambutas, V. & Beattie, D. S. *Biochim. biophys. Acta* **893**, 69–74 (1987).
16. Chow, W. S., Wagner, G. & Hope, A. B. *Austr. J. Pl. Physiol.* **3**, 853–861 (1976).
17. Hind, G., Nakatani, H. Y. & Izawa, S. *Proc. natn. Acad. Sci. U.S.A.* **71**, 1484–1488 (1974).

The maternal mRNA Vg1 is correctly localized following injection into *Xenopus* oocytes

Joel K. Yisraeli & D. A. Melton

Department of Biochemistry and Molecular Biology, Harvard University, Cambridge, Massachusetts 02138, USA

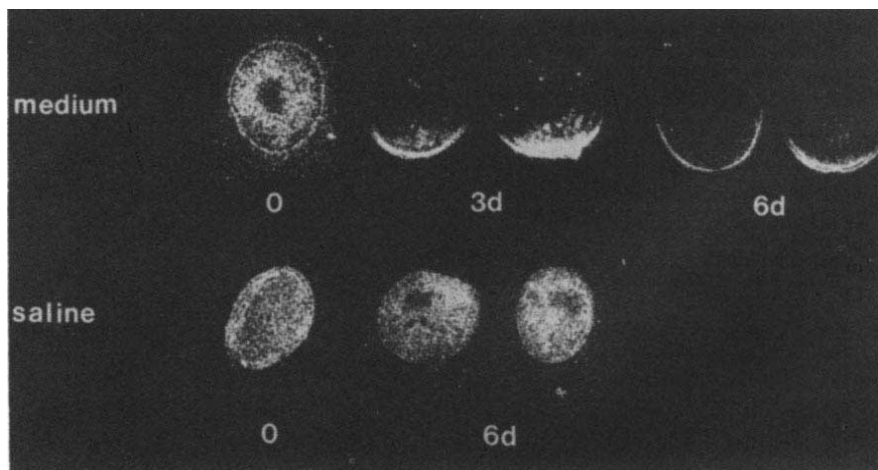
The animal and vegetal ends of *Xenopus* oocytes have distinctly different developmental fates. At the molecular level, several maternal mRNAs have been isolated that are localized to either the animal or vegetal hemisphere¹. One of these mRNAs, Vg1, is distributed homogeneously throughout the cytoplasm of early-stage oocytes and gets localized during oogenesis to a tight shell at the vegetal cortex of middle and late-stage oocytes². We have

used an *in vitro* culture system to demonstrate that exogenous Vg1 mRNA injected into middle-stage, but not late-stage, oocytes gets localized in a similar fashion to the endogenous message. Furthermore, translation of Vg1 mRNA is not required for the localization of the message itself. These results show that the information necessary to interpret the animal-vegetal polarity in oocytes is present in the naked mRNA transcript.

To study the Vg1 mRNA localization process, it was first necessary to obtain an *in vitro* culture system in which endogenous Vg1 mRNA is localized. Previous studies showed that endogenous Vg1 mRNA begins to localize in mid-stage oocytes (stage III, 0.5–0.6 mm in diameter²). We cultured stage III oocytes *in vitro* in a medium that contains vitellogenin³, and observed morphological changes associated with oocyte growth and development. In addition, *in situ* hybridizations show that the endogenous Vg1 mRNA is uniformly distributed in the cytoplasm at the beginning of the culture period, and is significantly concentrated in the vegetal hemisphere after only three

Fig. 1 Localization of endogenous Vg1 mRNA in cultured oocytes. Stage III albino oocytes (0.55–0.6 mm in diameter) were cultured in either medium supplemented with 10% vitellogenin-containing frog serum ('medium')^{3,4} or in 1× modified Barth's saline-HEPES (MBSH 'saline') for the indicated times. These dark-field photographs show the location of Vg1 transcripts as revealed by *in situ* hybridization with an antisense Vg1 RNA probe. Albino oocytes were used because the pigment granules in wild-type oocytes are almost indistinguishable from white autoradiographic silver grains. The nucleus or germinal vesicle (gv) appears as a dark circle in some sections. Here, as throughout the paper, whenever albino oocytes are used the sections are oriented with the putative vegetal pole down. In these middle-stage oocytes, the gv has not yet migrated to the animal hemisphere, making orientation of these albino oocytes difficult. On average the oocytes in medium grew in diameter from 0.55 mm at time 0 to 0.7 mm by day 6; those in saline showed a negligible increase. The ring of grains around the sections shown for time 0 results from an emulsion stretching artefact and is easily distinguished from hybridization when observed under the microscope.

Methods. Dissected oocytes (0.55 mm in diameter) were cultured in a 96-well plate at 20 °C in a humidified chamber in either 1× MBSH or 0.5× Leibowitz medium supplemented with glutamine, insulin, nystatin, gentamycin and 10% vitellogenin-containing frog serum^{3,4}. Blood containing serum rich in vitellogenin was isolated from female frogs injected three weeks earlier with 0.4 ml of 10 mg ml⁻¹ estradiol (Sigma) suspended in 1,3-propanediol³. At the end of the incubation, oocytes were fixed, sectioned, and hybridized² using a T7-synthesized [³²P]labelled, antisense Vg1 RNA probe containing the entire coding sequence of Vg1. Exposure time was five days to two weeks.



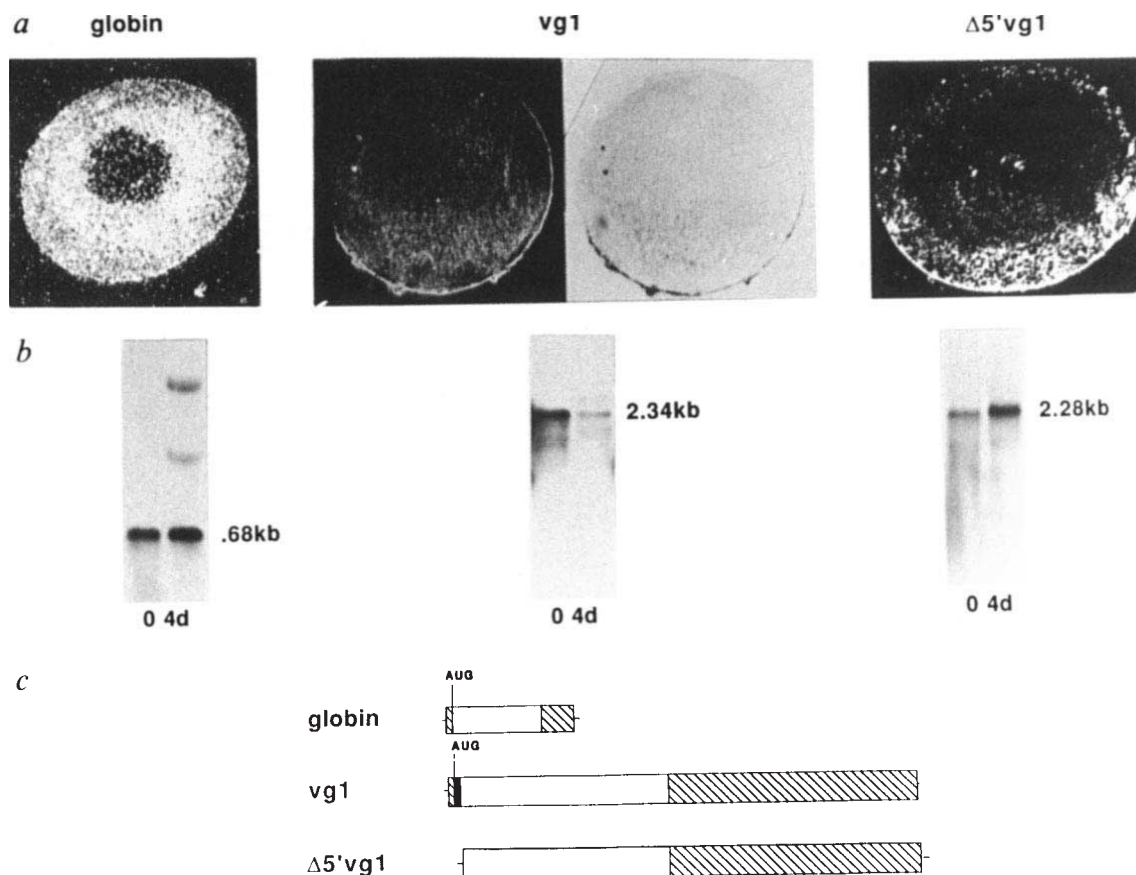


Fig. 2 Localization and stability of injected, labelled Vg1 mRNA. *a*, Late-stage III/early-stage IV oocytes (0.6–0.75 mm in diameter) were injected with synthetic, capped, [32 P]labelled, *Xenopus* β -globin (globin) mRNA, full-length Vg1 mRNA (Vg1), or Vg1 mRNA lacking 62 nucleotides at its 5' end ($\Delta 5'$ Vg1; see ref. 13) and cultured in growth medium for four days. For the Vg1 mRNA injection, both a dark-field (left) and bright-field (right) photograph are shown. The small amount of signal seen in the cytoplasm next to the cortical localization does not change with longer incubation. *b*, The stability of the injected, radioactive RNA described in *a*, was monitored by electrophoresis of the RNA through a formaldehyde-agarose gel. 0 represents the RNA extracted immediately after injection; 4d represents RNA from oocytes after four days of culture in growth medium. The sizes, in kilobases (kb), of the injected transcripts are indicated next to the bands on the gels. Two oocytes' worth of RNA was run in each lane of the globin and Vg1-injected oocyte RNA. For the $\Delta 5'$ Vg1-injected oocytes, only one oocyte's worth of RNA was run in lane 0, as opposed to three oocytes' worth in lane 4d. The extra, faint bands in the RNA extracted from the cultured oocytes (especially noticeable in the globin-injected oocytes, lane 4d) are rRNAs which have incorporated labelled UTP from degraded, injected RNAs. *c*, A schematic representation of the globin, Vg1, and $\Delta 5'$ Vg1 transcripts, with the hatched boxes representing untranslated regions, open boxes representing coding regions, the black box representing the putative signal sequence in Vg1 which is missing in $\Delta 5'$ Vg1 (as well as sequences further upstream; see ref. 13), and AUG indicating the position of the start codon.

Methods. Sp6 transcripts were synthesized from pSP64X β m (ref. 26) linearized with *Pst*I, pSP72Vg1 (ref. 13) linearized with *Eco*RI, or Gem3vg1.1 linearized with *Eco*RI (the 2.3 kb cDNA clone in ref. 13) in the presence of a cap analogue (GpppG) and [32 P]UTP to a specific activity of 3×10^7 c.p.m. μ g $^{-1}$ (ref. 8). Approximately 1 ng of RNA in 20 nl of sterile water was injected into the middle of each oocyte (see Fig. 3, time 0).

days of incubation (Fig. 1). By six days in culture, the characteristic tight crescent of Vg1 mRNA is observed. Thus, the localization of endogenous Vg1 mRNA can be obtained *in vitro* and these cultured oocytes are ideal recipients for injection experiments with exogenous Vg1 mRNA.

Notably, oocytes maintained in a standard saline solution do not grow or show any localization of Vg1 mRNA, although they remain viable. In fact, oocytes grown in vitellogenin-containing medium *in vitro* and those grown *in vivo* show a good correlation between oocyte size and the extent of Vg1 mRNA localization, even though the growth rate *in vitro*^{3,4} is faster than that *in vivo*⁵. Oocytes under 0.55 mm in diameter show a uniform Vg1 mRNA distribution. In 0.6 mm diameter oocytes, localization is in progress and is completed by the time the oocytes reach 0.7 mm in diameter (Fig. 1; J.K.Y., unpublished results). Significantly, yolk platelets in oocytes of this size are still uniformly distributed throughout the cytoplasm^{2,6}.

The rapid and striking localization of Vg1 mRNA is in marked

contrast to the vast majority of mRNAs in the oocyte which are distributed uniformly throughout the cytoplasm at all stages of oogenesis^{1,2,7}. This localization, coupled with the fact that the amount of Vg1 mRNA remains constant from stage II to stage VI oocytes, suggests that there is an active mechanism for translocating certain mRNAs such as Vg1². To test this hypothesis, Vg1 and globin mRNA were synthesized *in vitro*⁸ and injected into stage III oocytes. Injected Vg1 mRNA is localized in a pattern that is virtually indistinguishable from the endogenous Vg1 mRNA (Fig. 2a). This tight cortical localization seems to extend symmetrically along the entire vegetal cortex, as is seen with endogenous Vg1 mRNA (Fig. 1 and data not shown). Injected globin mRNA, by contrast, is distributed uniformly throughout the cytoplasm in oocytes cultured in parallel (Fig. 2a).

In addition to the final distribution of the injected Vg1 mRNA, its progressive movement also mirrors the localization of the endogenous message (Fig. 3). Both endogenous and exogenous

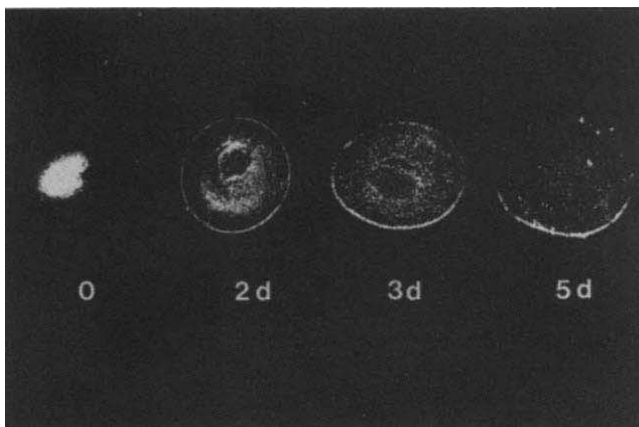


Fig. 3 Time course of the localization of the injected Vg1 RNA. Injection of labelled Vg1 RNA was as described in Fig. 2, with samples fixed and sectioned during culturing at the times indicated. The grains seen in the section at time 0 are in the middle of the oocyte, just to the left of the gv as oriented here. (The boundaries of the oocyte are not visible in dark-field.) The ring of grains surrounding the oocyte shown after two days in culture seems to be a result of the stretching of the emulsion (see Fig. 1).

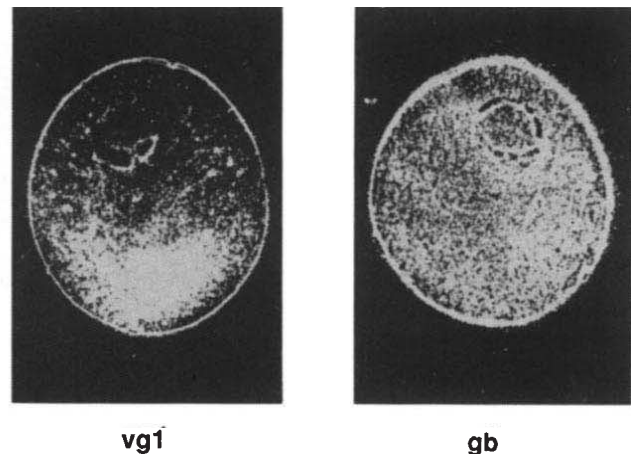


Fig. 4 Localization of injected RNA in late stage oocytes. Stage V/VI wild type oocytes were injected subequatorially with either labelled Vg1 (Vg1) or globin (gb) RNA and cultured for five days, as described in Fig. 2. Wild-type oocytes of this stage have a darkly pigmented animal hemisphere, but pigment grains are also visible in the vegetal hemisphere when the oocytes are sectioned and viewed in dark-field. There is little or no detectable signal at the cortex in either the Vg1-injected or globin-injected oocytes. In Vg1-injected oocytes, however, almost all of the grains appear in the vegetal hemisphere, whereas globin RNA appears to be fairly uniformly distributed throughout the oocyte. This pattern remains essentially unchanged with longer incubations.

Vg1 mRNA first become concentrated in one hemisphere and then begin to accumulate at the cortex (compare Figs 1 and 3).

Although localization of mRNA within a cell is poorly understood, more is known about the intracellular targeting of proteins (for reviews, see refs 9, 10). In particular, recognition of a nascent signal polypeptide at the N terminus of secretory proteins is an important step in protein translocation^{11,12}. Vg1 mRNA encodes a transforming growth factor (TGF) β -like molecule with a putative signal sequence at its 5' end and a peptidase cleavage site preceding the region of TGF β homology¹³. It was therefore of interest to know whether injected Vg1 mRNA needs to be translated to be localized.

A Vg1 transcript lacking the ribosome-binding site and the start codon as well as the signal sequence was constructed by deleting 62 nucleotides from the 5' end of the mRNA (Fig. 2c). When this 5' deleted mRNA is injected into stage III oocytes, it demonstrates a similar degree of localization as does the complete Vg1 mRNA (Fig. 2a). We cannot absolutely rule out the possibility of some translation at downstream start codons in the altered mRNA, but as the mRNA lacks its normal start codon and the entire signal sequence, these results provide evidence against any model in which localization of the mRNA would result from the recognition and/or targeting of a nascent Vg1 peptide chain. Instead, the results point to a localization process that acts through specific recognition of the mRNA (or mRNA-protein complex) itself.

The stability of the injected mRNAs was assayed by gel electrophoresis of RNA extracted from injected oocytes at the end of the culture period (Fig. 2b). Typically, about 25–50% of injected Vg1 and 5' deleted Vg1 mRNA remains intact after four days, and 50–80% of injected globin mRNA is stable. The globin RNA is unusually stable in injected oocytes, probably because of its 3' poly(A)¹⁴; the synthetic Vg1 mRNA may be less stable because it lacks poly(A) at its 3' end. Reincorporation of labelled nucleotides from degraded, injected RNA was only detected in ribosomal RNAs (see Fig. 2b), and most of the labelled nucleotides were found in the culture medium. Thus, reincorporated label does not contribute to the localized signal seen from injected RNA. In fact, only a faint background of uniformly distributed grains is detected when [³²P]UTP alone is injected into oocytes (data not shown).

It is formally possible that localized degradation is partly responsible for the observed distribution of Vg1 RNA, but the stability of the injected RNA does not support this view. Furthermore, no difference in stability is detected in Vg1 mRNA injected into oocytes cultured in growth medium (where localization is observed) or cultured in saline (where no localization is observed) (data not shown). Alternatively, passive diffusion coupled with increased affinity of Vg1 RNA to the vegetal cortex could result in the localization of Vg1 RNA. Because the accumulation of Vg1 RNA along the cortex is coordinated, however, with the disappearance of Vg1 RNA from the cytoplasm in an animal to vegetal direction (Fig. 1 and see below; ref. 2), we favour a model for localization in which Vg1 RNA is actively translocated to the vegetal end of the oocyte.

A diffuse signal over most of the vegetal hemisphere is discernible in some of the cultured, Vg1-injected stage III oocytes (Fig. 2a) and may be the result of injecting a relatively large amount of mRNA (1 ng). A similar result is obtained when late-stage oocytes are injected (Fig. 4), although in this case no cortical localization of Vg1 mRNA is detected. These results suggest that the localization may consist of two processes, one that transports the RNA to the vegetal hemisphere and another that anchors it along the cortex.

A growing number of RNA species have been reported to be asymmetrically distributed in both mononucleated cells^{1,6,15,16} and syncytial muscle¹⁷. In the early development of the *Drosophila* embryo, localized *bicoid* mRNA is essential for the proper formation of the anterior body plan^{18–21}, and recent studies have identified the RNA signals involved in *bicoid* mRNA sequestration²⁷. The localization of Vg1 mRNA to the endoderm and its sequence similarity to TGF β (an inducer of mesoderm in animal cap explants^{22,23}), suggest that Vg1 may play a role in specifying cell fates in early frog development. For all these reasons, it will be of interest to extend these studies to define a *cis*-acting sequence responsible for the correct localization of Vg1 mRNA. In other experiments, detergent-insoluble extracts of oocytes²⁴ preferentially contain 80–90% of all Vg1 mRNA in the cell, as opposed to most other mRNAs, which are in the soluble fraction (S. Sokol, M. R. Rebagliati and

D.A.M., unpublished observations²⁸). This cytoskeletal association of the Vg1 mRNA is reminiscent of RNA-cytoskeletal associations reported in other systems^{24,25}. By specifically interfering with different cytoskeletal elements, and by identifying cytoplasmic components that specifically recognize Vg1 mRNA, it should be possible to dissect the localization process and to understand better how polarity is established and interpreted.

We would like to thank all of our colleagues whose constructive criticisms have greatly improved this manuscript: Anne Ephrussi, Charles Jennings, Paul Krieg, Ariel Ruiz i Altaba, Sergei Sokol, David Tannahill and Malcolm Whitman. This work was supported by a Chaim Weizmann Postdoctoral Fellowship (J.K.Y.) and an NIH grant (D.A.M.).

Received 22 August; accepted 9 November 1988.

1. Rebagliati, M. R., Weeks, D. L., Harvey, R. P. & Melton, D. A. *Cell* **42**, 769-777 (1986).
2. Melton, D. A. *Nature* **328**, 80-82 (1987).
3. Wallace, R. A., Misulovin, Z. & Wiley, H. S. *Reprod. Nutr. Dev.* **20**, 699-708 (1980).
4. Wallace, R. A., Misulovin, Z., Jared, C. W. & Wiley, H. S. *Gamete Res.* **1**, 269-280 (1978).
5. Keem, K., Smith, L. D., Wallace, R. A. & Wolf, D. *Gamete Res.* **2**, 125-135 (1979).
6. Danilchik, M. V. & Gerhart, J. C. *Dev. Biol.* **122**, 101-112 (1987).
7. King, M. L. & Barklis, E. *Dev. Biol.* **112**, 203-212 (1985).
8. Krieg, P. A. & Melton, D. A. *Meth. Enzym.* **155**, 397-415 (1987).
9. Walter, P., Gilmore, R. & Blobel, G. *Cell* **38**, 5-8 (1984).
10. Colman, A. & Robinson, C. *Cell* **46**, 321-322 (1986).
11. Hortin, G. & Boime, I. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1356 (1980).
12. Walter, P. & Blobel, G. *J. Cell Biol.* **91**, 545 (1981).
13. Weeks, D. L. & Melton, D. A. *Cell* **51**, 861-867 (1987).
14. Drummond, D. R., Armstrong, J. & Coleman, A. *Nucleic Acids Res.* **13**, 7375-7394 (1985).
15. Lawrence, J. B. & Singer, R. H. *Cell* **45**, 407-415 (1986).
16. Jeffery, W. R., Tomlinson, C. R. & Brodeur, R. D. *Dev. Biol.* **99**, 408-417 (1983).
17. Fontaine, B., Sasson, D., Buckingham, M. & Changeux, J.-P. *EMBO J.* **7**, 603-609 (1988).
18. Frigerio, G., Burri, M., Boppo, D., Baumgartner, S. & Noll, M. *Cell* **47**, 735-746 (1986).
19. Nusslein-Volhard, C., Frohnhofer, H. G. & Lehmann, R. *Science* **238**, 1675-1681 (1987).
20. Driever, W. & Nusslein-Vollard, C. *Cell* **54**, 83-93 (1988).
21. Driever, W. & Nusslein-Vollard, C. *Cell* **54**, 95-104 (1988).
22. Kimmelman, D. & Kirschner, M. *Cell* **51**, 869-877 (1987).
23. Rosa, F. *et al. Science* **239**, 783-785 (1988).
24. Jeffery, W. R. *Dev. Biol.* **103**, 482-492 (1984).
25. Davis, L., Banker, G. A. & Steward, O. *Nature* **330**, 477-479 (1987).
26. Krieg, P. A. & Melton, D. A. *Nucleic Acids Res.* **12**, 7057-7071 (1984).
27. Macdonald, P. M. & Struhl, G. *Nature* **336**, 595-598 (1988).
28. Pondel, M. D. & King, M. L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7612-7616 (1988).

Cis-acting sequences responsible for anterior localization of *bicoid* mRNA in *Drosophila* embryos

Paul M. Macdonald & Gary Struhl

Howard Hughes Medical Institute, Center for Neurobiology and Behavior, Columbia University College of Physicians and Surgeons, 722 W. 168 St, New York, N.Y. 10032, USA

The anterior body pattern of *Drosophila melanogaster* is specified in large part by the protein product of the *bicoid* (*bcd*) gene which functions as a graded morphogen with its peak of expression at the anterior pole of the embryo¹⁻⁵. Formation of the gradient is dependent on prior localization of *bcd* messenger RNA at the anterior pole of the egg cell during oogenesis^{6,7}. Here we demonstrate that a discrete portion of the *bcd* mRNA is necessary for anterior localization of the *bcd* transcript and is sufficient to cause localization of heterologous transcripts. The sequences responsible for localization appear to span an interval of about 625 base pairs in the 3' untranslated portion of the *bcd* mRNA and to include regions capable of forming extensive secondary structure. Transcripts from *bcd* are synthesized predominantly, if not exclusively, in the nurse cells and then transported to the oocyte by connections at the prospective anterior pole^{6,7}. Our findings support the proposal² that *bcd* transcripts are selectively recognized and trapped as they enter the anterior tip of the oocyte, and suggest that this localization process is mediated by anchored sequence-specific receptors in the oocyte cytoplasm.

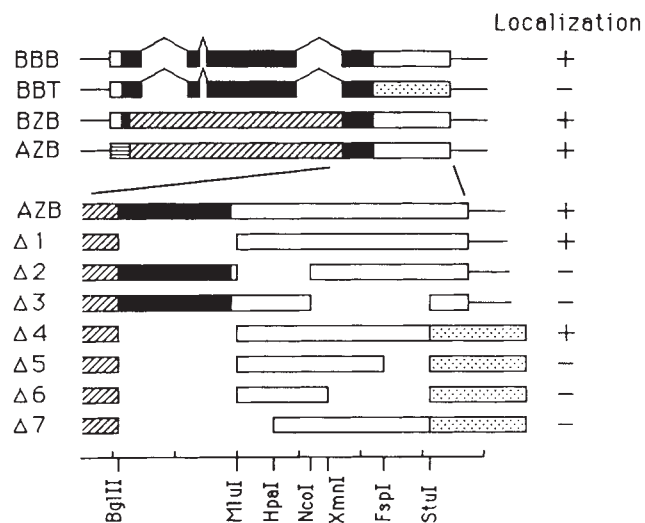
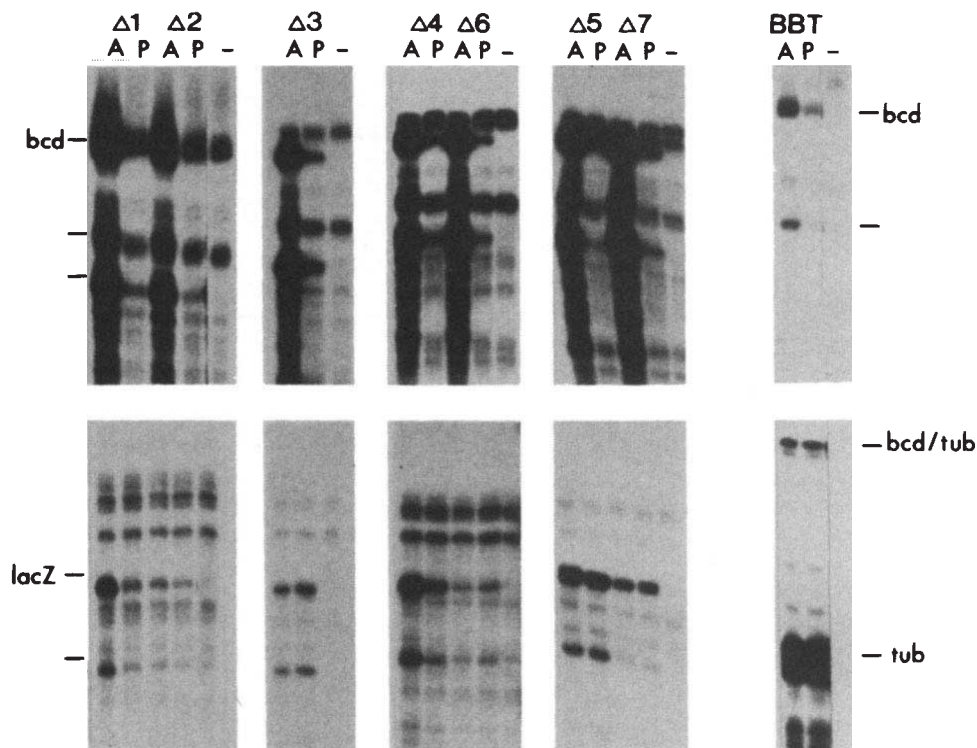


Fig. 1 Mapping the *bcd* mRNA localization signal. The *bcd* gene and relevant portions of the plasmids constructed to map the *bcd* mRNA localization signal are shown in schematic form, together with an indication of the ability of each to confer localization to its mRNA in embryos (see Fig. 2). The *bcd* gene is shown (not to scale) in the context of an 8 kilobase (kb) *EcoRI* fragment which rescues *bcd*⁻ flies to wild type when inserted by P-element transformation (construct BBB; see also ref. 7). Exon sequences are boxed, with all but the 5' and 3' noncoding regions filled in. In BZB, *lacZ* sequences (diagonal hatching) replace most of the *bcd* coding region. In AZB, all *bcd* 5' regions are replaced by the *Adh* promoter and 5' flanking sequences (horizontal hatching). The only *bcd* sequences remaining are from the 3' region, which is expanded and presented in scale to outline the changes introduced in Δ1-Δ7 (deletions are shown as empty spaces). Δ1-Δ3 retain the *bcd* polyadenylation signal and downstream sequences, whereas in Δ4-Δ7 those regions are replaced with the polyadenylation signal present in a segment of the *atub84B* gene¹⁴ (shown as a stippled box). A partial restriction map of the region is presented at the bottom, showing only the sites used to create the deletions. Scale increments are 200 base pairs (bp) each.

Methods. Transgenic flies were generated by standard techniques^{8,9}. All plasmids were constructed in the Carnegie 20 transformation vector, variously modified to include *XbaI*, *NotI*, and *Asp718* cloning sites in the polylinker. The genomic *bcd* *EcoRI* fragment (BBB) was derived from a phage clone²⁰ from the Antennapedia complex (provided by M. Scott); transcript structure and sequence were established by standard techniques (P.M.M., unpublished results; ref. 7). BZB was constructed in several steps to yield the following: *bcd* sequences extending from the 5' *EcoRI* site to a *BamHI* linker inserted just downstream from amino acid 5 in the coding sequence; *lacZ* sequences extending from a *BamHI* site near the amino terminus to just downstream from the stop codon; *bcd* 3' sequences continuing from a natural *BglII* site at the 5' end of the final exon to the 3' *EcoRI* site of BBB. The *bcd* DNA deleted in this construct appears to include an element required for the normal high level of oocyte expression, because transcripts accumulate to less than a tenth of the normal amount. AZB was derived from BZB by replacing the 5' *bcd* and *lacZ* sequences with an *Adh/lacZ* hybrid gene in which *Adh* is fused in-frame to *lacZ* immediately after the ATG (V. Corbin and T. Maniatis, personal communication). The *Adh* promoter fragment extends to an *XbaI* site at -600 bp relative to the distal transcription start site, and in other constructs is sufficient to direct high levels of *lacZ* transcript accumulation in oocytes (V. Corbin and T. Maniatis, personal communication). However, when present with the 3' *bcd* sequences, as shown for AZB, only low levels of *lacZ* transcript are detected. For Δ4-Δ7, the 5' DNA was altered, primarily by adding an additional 4.4 kb of 5' *Adh* flanking sequences, in an unsuccessful attempt to increase the level of expression. Otherwise, Δ1-Δ7 were derived from AZB by using the restriction sites shown on the map to delete various portions of the *bcd* 3' region. For Δ4-Δ7, the common deletion of all *bcd* DNA 3' to the *StuI* site removes the polyadenylation signal which was replaced by a fragment of the *atub84B* gene (nucleotides 1,960-2,770 in ref. 14) encoding the 3' non-translated flanking portion of the mRNA and including the polyadenylation signal.

Fig. 2 Transcript distributions in early embryos. Relative levels of various hybrid transcripts and the endogenous *bcd* transcripts from the anterior and posterior halves of early embryos were determined by standard RNase protection assays²¹. Note that RNAs protect both full-length complementary *bcd* and *lacZ* probes as well as one or more smaller probe fragments (each indicated by a dash). Note also that the RNAs in each experiment were obtained from flies carrying 1–2 copies of a single *P(ry⁺ AZBΔ)* element; hence the level of expression of the *lacZ*-containing transcripts varies over a few-fold range for the different transgenes tested. For constructs Δ1–Δ7 (see Fig. 1), the upper panels show the levels of endogenous *bcd* mRNA in anterior (A) and posterior (P) samples, and the lower panels show the levels of *lacZ* mRNA (from the hybrid genes) in the same RNA samples; lanes labelled with a dash are the control samples to which tRNA, but no embryo RNA, was added: these indicate the presence of probe-dependent background bands (note that at least one of the three protected *bcd* fragments run separately from these background bands in each experiment). For each transgenic line most of the endogenous *bcd* mRNA is found in the anterior fraction. In contrast, the *lacZ* mRNA is asymmetrically distributed only in Δ1 and Δ4, whereas in Δ2, Δ3, Δ5, Δ6 and Δ7 similar levels are present in both A and P fractions. Each sample was also tested with a probe specific for the *rp49* mRNA to confirm that A and P fractions contained similar amounts of total RNA (not shown). For BBT, the upper panel shows the relative levels of *bcd* mRNA in A and P fractions. Because the *bcd* probe used in this assay detects the same portion of both the localized endogenous *bcd* and nonlocalized BBT mRNAs, there is a significant signal attributable to *bcd/tub* hybrid transcripts in the P sample. In the lower panel, the same RNA samples were tested with a hybrid *bcd/αtub48B* probe. The band corresponding to the hybrid mRNA is labelled *bcd/tub* and is present at similar levels in both A and P samples; the band which is due to the endogenous *αtub84B* mRNA is labelled *tub* and provides a control showing that similar amounts of RNA were present in A and P fractions.



For each transgenic line most of the endogenous *bcd* mRNA is found in the anterior fraction. In contrast, the *lacZ* mRNA is asymmetrically distributed only in Δ1 and Δ4, whereas in Δ2, Δ3, Δ5, Δ6 and Δ7 similar levels are present in both A and P fractions. Each sample was also tested with a probe specific for the *rp49* mRNA to confirm that A and P fractions contained similar amounts of total RNA (not shown). For BBT, the upper panel shows the relative levels of *bcd* mRNA in A and P fractions. Because the *bcd* probe used in this assay detects the same portion of both the localized endogenous *bcd* and nonlocalized BBT mRNAs, there is a significant signal attributable to *bcd/tub* hybrid transcripts in the P sample. In the lower panel, the same RNA samples were tested with a hybrid *bcd/αtub48B* probe. The band corresponding to the hybrid mRNA is labelled *bcd/tub* and is present at similar levels in both A and P samples; the band which is due to the endogenous *αtub84B* mRNA is labelled *tub* and provides a control showing that similar amounts of RNA were present in A and P fractions.

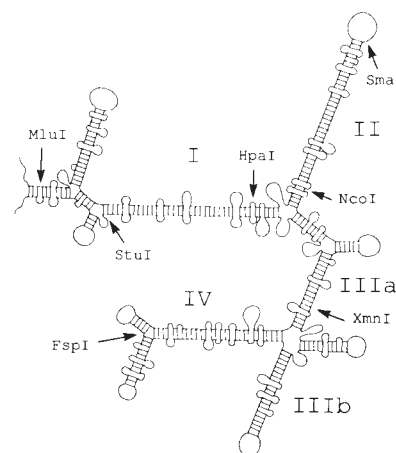
Methods. 0–1-hour-old embryos were collected on yeasted agar plates and mounted side-by-side on double-sided adhesive tape on a microscope slide. The embryos were then immersed in 3S Voltalef oil, the slide placed on a metal block sitting on dry ice and the embryos, once frozen, bisected with a razor blade (prechilled in liquid nitrogen). A and P embryo halves were then transferred to microfuge tubes containing RNA extraction buffer and the RNA was purified as described previously²². Templates for synthesis²³ of antisense RNA probes were the following: *bcd*: plasmid p971, which consists of a 400 bp *PstI*–*HincII* restriction fragment (from within the third *bcd* exon) cloned into the *PstI* and *SmaI* sites of pGEM1 was cut with *HindIII* (adjacent to *PstI* in the polylinker) and transcribed with T7 RNA polymerase. All 400 nucleotides from the *bcd* DNA will be protected by *bcd* mRNA. *lacZ*: plasmid p727, which consists of a 1,343 *AvaI* restriction fragment from the *lacZ* gene cloned into pGEM1 after blunting and addition of *HindIII* linkers, was cut with *MluI* and transcribed with T7 RNA polymerase to produce a probe of which 274 nucleotides will be protected by the *lacZ* component of all constructs. *rp49*: plasmid p720, which consists of a 253 bp *EcoRI*–*ClaI* genomic *rp49* restriction fragment¹⁰ cloned into the *EcoRI* and *AccI* sites of pSP64, was cut with *HindIII* (near *AccI* in the polylinker) and transcribed with SP6 RNA polymerase to yield a probe of which 81 nucleotides will be protected by the 5' portion of the *rp49* mRNA. *bcd/tub*: plasmid p1166, which consists of a 262 bp *HpaI* fragment from BBT (Fig. 1; 44 bp of *bcd* 3' noncoding region, 18 bp of polylinker sequence and 200 bp of *αtub84B* DNA) cloned into pGEM1, was cut with *HindIII* (in the polylinker) and transcribed with T7 RNA polymerase to produce a probe in which 262 nucleotides will be protected by the BBT mRNA, and 200 nucleotides by the *αtub84B* mRNA. The RNase protection assays were done as described previously²², except that each hybridization included 20 μg carrier tRNA. For Δ1 and Δ2, 20% and 80% of each sample was used to probe for the *bcd* and *lacZ* mRNAs respectively. For the others in the Δ series, the proportions were 10% and 90%, and for BBT half of each sample was used for each probe. In general, the amount of RNA recovered in each A and P sample was about 5–10 μg.

To determine whether the *bcd* mRNA contains a *cis*-acting sequence necessary and sufficient for localizing the transcript at the anterior pole of the embryo, we have examined the antero-posterior distributions of transcripts derived from hybrid genes in which various portions of the *bcd* gene are replaced by analogous portions of other genes (see Fig. 1). These hybrid genes were introduced into the *Drosophila* genome by P-element mediated transformation^{8,9}. Early embryos obtained from transgenic mothers carrying these genes were cut into anterior and posterior halves and the RNAs from these halves assayed separately for transcripts from (1) the hybrid gene, (2) the endogenous *bcd* gene, which are localized, and (3) a ribosomal protein gene (*rp49*; ref. 10), which are not localized. These experiments identify a discrete region of the *bcd* transcript positioned 3' to

the coding sequence which is both necessary and sufficient for anterior localization of the *bcd* mRNA.

We began with an 8-kilobase segment of genomic DNA which contains a fully functional *bcd* gene (BBB in Fig. 1), as assayed by the ability of this segment to rescue the development of *bcd*[−] animals. A portion of the *bcd* gene containing most of the coding sequence was then replaced by the *lacZ* coding sequence to make BZB (Fig. 1); mRNA from BZB was localized in the anterior half of the embryo (data not shown), indicating that signals sufficient to localize the transcript were retained in this hybrid gene.

To determine which of the remaining *bcd* regions are responsible for localizing the BZB hybrid transcript, we first tested the importance of the promoter and 5' transcribed sequences by



The existence of a discrete *cis*-acting signal that is necessary and sufficient to localize *bcd* mRNA provides strong support for the proposal² that the transcripts are selectively recognized and trapped after they pass from the nurse cells by cytoplasmic bridges to the anterior tip of the growing oocyte. These results also argue for the existence of one or more receptors capable of recognizing and binding to the localization signal on the *bcd* transcript (or perhaps on *bcd*-associated ribonucleoprotein particles), and for the ability of these receptors to be bound to fixed structural components of the egg cell, perhaps through distinct anchoring molecules. Note that neither the receptors nor anchors needs to be localized within the oocyte because polarized entry

of the transcripts would provide the necessary asymmetry in the system. The only spatial restriction that this model imposes is that the trapping mechanism be functional only in the oocyte; otherwise *bcd* transcripts would be trapped before they could exit from the nurse cells.

Localization of *bcd* transcripts during *Drosophila* oogenesis appears to differ from that of *veg1* transcripts¹⁹ during *Xenopus* oogenesis. In the latter case, initially uniform transcripts become progressively localized within the oocyte, suggesting that mRNA localization depends on a prior localization of the presumed receptor or anchoring molecules within the egg. Nevertheless, the mechanisms by which these RNAs are selectively recognized and bound to subcellular components could be similar in both cases. In *Drosophila*, the initial cue for establishing the anterior patterning system may be the cellular organization of the nurse cell/oocyte complex: in particular the constraint that all of the synthetic products of the nurse cells must enter the oocyte at its anterior end¹⁸. This, together with a relatively simple mechanism for binding and anchoring a defined mRNA species to a fixed subcellular component suffices to create the necessary preconditions for forming the *bcd* morphogen gradient.

We thank Pat Micelli and Jane Albert-Hubbard for technical assistance, John Fisher for performing the computer analysis of potential RNA secondary structure, Jordi Casanova, Ken Howard, Robin Wharton and Tom Jessell for discussion,

Matthew Scott for the genomic clones of the *bcd* gene, Vicki Corbin and Janice Fischer for the *Adh* promoter fragments, Quang Yu for the *rp49* clone and Pieter Wensink for the *atub84A* clone. We also thank the Howard Hughes Medical Institute and the McKnight Foundation for financial support.

Received 13 September; accepted 8 November 1988.

1. Frohnhof, H. G. & Nusslein-Volhard, C. *Nature* **324**, 120-125 (1986).
2. Frohnhof, H. G. & Nusslein-Volhard, C. *Genes Dev.* **1**, 880-890 (1987).
3. Nusslein-Volhard, C., Frohnhof, H. G. & Lehmann, R. *Science* **238**, 1675-1681 (1987).
4. Driever, W. & Nusslein-Volhard, C. *Cell* **54**, 83-93 (1988).
5. Driever, W. & Nusslein-Volhard, C. *Cell* **54**, 95-104 (1988).
6. Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. *Cell* **47**, 735-746 (1986).
7. Berleth, T. *et al. EMBO J.* **7**, 1749-1756 (1988).
8. Spradling, A. C. & Rubin, G. M. *Science* **218**, 341-347 (1982).
9. Rubin, G. M. & Spradling, A. C. *Science* **218**, 348-353 (1982).
10. O'Connell, P. & Rosbash, M. *Nucleic Acids Res.* **12**, 5495-5513 (1984).
11. Goldberg, D. A., Posakony, J. W. & Maniatis, T. *Cell* **34**, 59-73 (1983).
12. Posakony, J. W., Fischer, J. A. & Maniatis, T. *Cold Spring Harbor Symp. quant. Biol.* **50**, 515-520 (1986).
13. Savakis, C., Ashburner, M. & Willis, J. H. *Dev. Biol.* **114**, 194-207 (1986).
14. Theurkauf, W. E., Baum, H., Bo, J. & Wensink, P. C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8477-8481 (1986).
15. Kozak, M. *J. Cell. Biol.* **107**, 1-7 (1988).
16. Leibold, E. A. & Munro, H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2171-2175 (1988).
17. Zuker, M. & Stiegler, P. *Nucleic Acids Res.* **9**, 133-148 (1981).
18. King, R. C. *Ovarian Development in Drosophila melanogaster* (Academic, New York, 1970).
19. Melton, D. A. *Nature* **328**, 80-82 (1987).
20. Scott, M. P. *et al. Cell* **35**, 763-776 (1983).
21. Zinn, K., DiMaio, D. & Maniatis, T. *Cell* **34**, 865-879 (1983).
22. Macdonald, P. M., Ingham, P. W. & Struhl, G. *Cell* **47**, 721-724 (1986).
23. Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
24. Freier, S. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 9373-9377 (1986).

***Drosophila* homoeotic genes encode transcriptional activators similar to mammalian OTF-2**

Markus Thali*, Michael M. Müller*,
Mauro DeLorenzi†, Patrick Matthias*
& Mariann Bienz†‡

* Arbeitsgruppe Walter Schaffner, Institut für Molekularbiologie II, Universität Zürich, Höggerberg, 8093 Zürich, Switzerland

† Zoologisches Institut, Universität Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland

Homoeotic genes in *Drosophila melanogaster* are active in spatially restricted metameric domains and control the morphogenesis of segment-specific features such as legs or wings within these domains¹. They exert their function, according to the 'selector gene' hypothesis², by regulating the expression of subordinate genes. Homoeotic genes also control their own expression³ and the expression of each other³⁻⁵. The proteins encoded by these genes contain a domain, called a homoeodomain, that is strongly conserved, and that shows homologies to proteins that bind DNA and regulate transcription^{6,7}. Homoeoproteins have been shown to bind specific DNA sequences⁸⁻¹⁰. We show here that the *Drosophila* homoeotic genes *Ultrabithorax* (*Ubx*) and *Abdominal-B* (*Abd-B*) code for proteins that are capable of activating transcription of reporter genes linked to specific *cis*-regulatory target sequences in transfected mammalian cells. Their activity, as well as their target specificity, is similar to that of a mammalian lymphoid-specific octamer transcription factor, OTF-2, which was recently found to contain a homoeodomain¹¹.

Eukaryotic transcriptional activators can function, as a rule, in heterologous eukaryotic cells by stimulating transcription of a co-transfected reporter gene bearing suitable *cis*-regulatory target sequences¹². One of these activators, the lymphoid-specific transcription factor OTF-2, which binds to the 'octamer' motif T-N-A-T-T-T-G-C-A-T, functions in transfected HeLa cells as a transcriptional activator on octamer-containing B-cell-specific

promoters¹¹. Sequence analysis revealed that this transcription factor contains a homoeodomain¹¹. We asked whether *Drosophila* homoeoproteins might be transcriptional activators similar to mammalian OTF-2 and whether they would be functional as activators by themselves after transfection into mammalian cells.

We tested the *Ubx* and *Abd-B* homoeoproteins for their potential as transcriptional activators in HeLa cells. We expressed these proteins from their complementary DNAs^{13,14} subcloned into an expression vector¹⁵. As a putative *cis*-regulatory target, we used a short sequence stretch (B) derived from the RNA leader region of the *Ubx* gene, as this had been shown to constitute a *Ubx* protein binding site (P. Beachy, M. Krasnow, E. Gavis and D. Hogness, personal communication; see also ref. 16) and to be required for correct spatial expression of a *Ubx*/ β -galactosidase (*Ubx*/ β -gal) fusion gene in transformed embryos (ref. 17; J. Müller and M.B., manuscript in preparation). As a second target, we used a tandem repeat as it naturally occurs in the *engrailed* gene¹⁰ (E) of the sequence T-C-A-A-T-T-A-A-A-T previously identified as a consensus sequence for homoeoprotein binding^{9,10}. The targets were cloned upstream of a globin gene TATA-box into OVEC¹⁸ vectors (OVEC B, OVEC E) with or without an SV40 enhancer downstream.

For comparison, we used a cDNA encoding mammalian OTF-2 subcloned in the same expression vector. As targets for OTF-2, we tested two synthetic B-cell-specific octamer-binding sites that cannot direct transcription in HeLa cells in the absence of OTF-2¹¹ (OVEC Octa1, OVEC Octa2).

To control for target specificity, we tested two further OVEC plasmids containing regulatory sequences that are recognized by other known transcription factors (OVEC 2SP1 bearing two SP1-binding sites¹⁹; OVEC 4GRE bearing four palindromes of glucocorticoid response elements¹⁵) and therefore should not respond to our *trans*-activators. We also constructed an OVEC plasmid containing a B substitution sequence (OVEC Bsub) that we had used to replace the homoeoprotein binding site in the *Ubx*/ β -gal gene to abolish its expression in transformed embryos (ref. 17; J. Müller and M.B., manuscript in preparation). All plasmids are depicted in Fig. 1.

We co-transfected HeLa cells with cDNA-containing expression vectors and OVEC reporter plasmids in all possible combinations. Reporter gene expression was measured by SP6-

‡ To whom correspondence should be addressed.

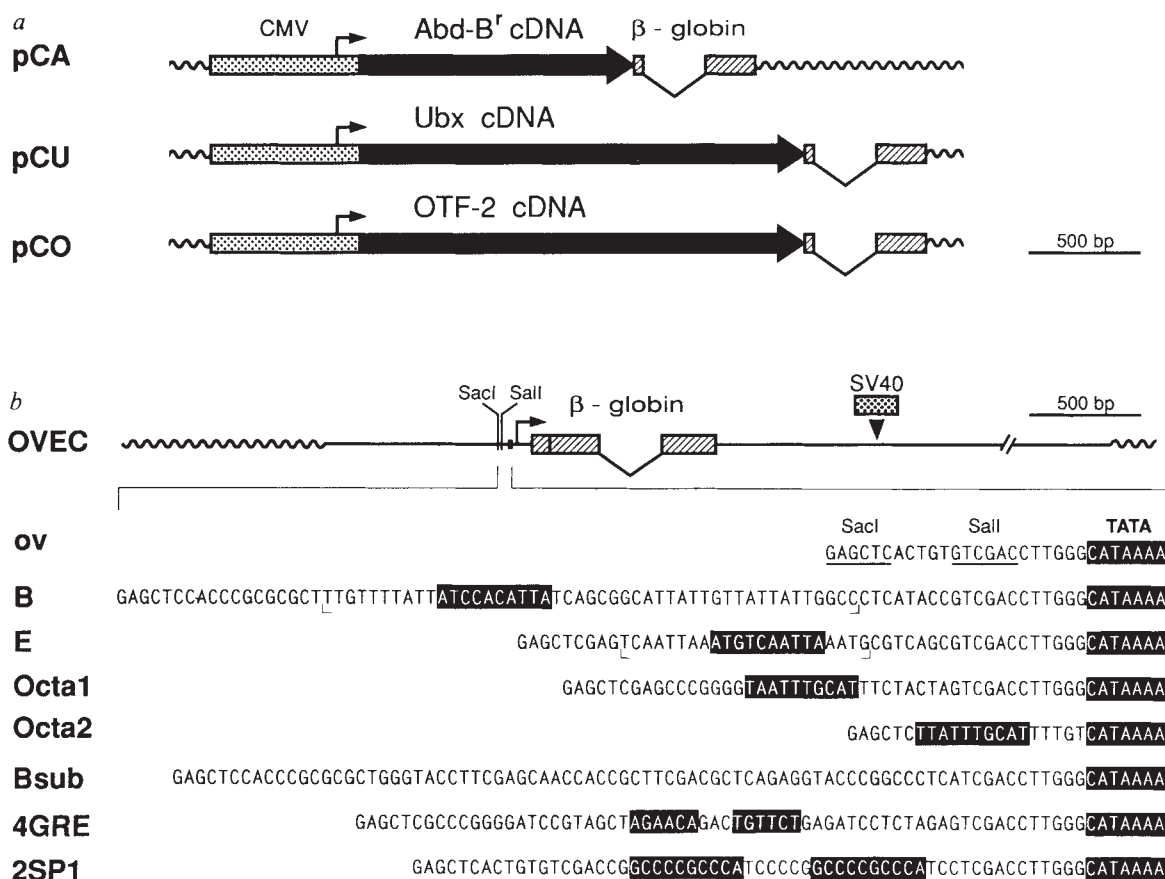


Fig. 1 Plasmids containing *trans*-activators and *cis*-regulatory target sequences. **a**, *Trans*-activators. Maps of three expression plasmids containing cDNAs (black) for *Abd-B^r* (pCA), *Ubx* (pCU) or *OTF-2* (pCO) with their translation start and stop codons, linked to a cytomegalovirus (CMV) promoter (stippled; pCO contains additional thymidine kinase leader sequences¹¹) and to β -globin sequences (solid line; coding part hatched) including the second intron and polyadenylation signal¹⁵. An SV40 origin of replication is indicated by a black circle. bp, base pairs. **b**, Target-bearing OVEC plasmids. A map of OVEC¹⁸ containing, as a reporter gene, β -globin gene sequences (solid line; coding part hatched; triangle indicates deletion of the first intron) including the TATA box (little black square) as the only functional upstream promoter element. Immediately upstream, a unique *SacI* and *Sall* site for insertion of various target sequences as indicated below. **ov**, OVEC plasmid without insert, TATA-box highlighted, cloning sites underlined; **B**, functional homeoprotein binding site from the *Ubx* gene (ref. 17; J. Müller and M.B., manuscript in preparation), footprint area¹⁶ in brackets; **E**, tandem duplication, in brackets, of a homeoprotein-binding consensus sequence^{9,10} (boxed); **Octa1**, **Octa2**, octamer motifs (boxed) similar to those described¹¹; **Bsub**, non-functional substitution sequence¹⁷ of the B element; **4GRE**, palindromes of four glucocorticoid response elements as previously published¹⁵ only one of which is depicted (in brackets; palindrome boxed); **2SP1**, two SP1 binding sites (boxed) as previously published¹⁹. A sequence motif in **B** and **E** similar to the octamer motif (see Fig. 3) is framed by dotted lines. The position of an optional SV40 enhancer (see text) is given. Transcription starts marked by arrows. Wavy lines, vector sequences.

Methods. To generate pCA, an *MluI/BglII* 1.2 kilobase (kb) fragment from an *Abd-B^r* cDNA clone¹⁴ was inserted, via a subcloning step into *SmaI/BamHI*-cut pUC19, as a filled-in *EcoRI/XbaI* fragment into a vector modified from pSTC (ref. 15) that lacks thymidine kinase leader sequences (S. Rusconi, unpublished) and that contains *BamHI* and *XbaI* as the only cloning sites. To generate pCU, an *XhoII/HincII*-cut 1.9 kb fragment^{29,31} from a '1S type' *Ubx* cDNA clone¹³ was inserted similarly via a *BamHI/SmaI*-cut pUC intermediate between the *BamHI* and the *XbaI* site of the modified pSTC vector. pCO is a variant from pOEVI(+) (ref. 11), but equally active in *trans*-activation; to generate pCO, the OTF-2 cDNA fragment in pOEVI(+) was substituted by the OTF-2 cDNA fragment from λ OB-B (ref. 11). For OVEC **B**, a short genomic filled-in *BssHII*/filled-in *DdeI* fragment from the *Ubx* gene (positions +37/+93; ref. 31) was inserted, via a bluescript subcloning step, into *SacI/Sall*-cut OVEC. For OVEC **Bsub**, a *BssHII*/filled-in *DdeI* fragment from the substitution plasmid -626sub (ref. 16) was inserted into *BssHII*/filled-in *Sall*-cut OVEC **B** to replace the **B** sequence. The OVEC **E** plasmid was constructed with a double-stranded synthetic oligomer. The inserts of all these plasmids were sequenced for confirmation. The OVEC **2B** and **2E** plasmid were generated by *XhoI/Sall* duplication¹⁸. The plasmids bearing octamer targets correspond to plasmids Octa1 (ref. 11) and +5junc⁻ (Octa2; P. M., unpublished). Other OVEC plasmids as previously published^{15,19}.

RNase protection assays and compared to expression of a co-transfected reference gene (OVEC-REF; ref. 18), a globin gene linked at its TATA-box to the SV40 enhancer. The results are shown in Fig. 2a. In the absence of a functional co-transfected *trans*-activator, our targets resulted in low or undetectable reporter gene expression, with the exception of OVEC 2SP1 whose expression reflects endogenous SP1 activity¹⁹. However, co-transfection of the *Abd-B^r* cDNA plasmid pCA (lanes marked a) induced very strong reporter gene expression via the targets

B and **E** (up to 144-fold), but not via the control targets 4GRE or **Bsub**. We observed some enhancement of OVEC 2SP1 expression, exaggerated by the low level of reference plasmid expression in the lane marked a (all 'a' lanes show reduced reference plasmid expression, probably due to a squelching effect¹² by *Abd-B^r*). A fortuitous homeoprotein binding site in the upstream β -globin sequences might be the cause for this distance effect, which is only mediated if upstream elements adjacent to the TATA-box are occupied with functional tran-

Fig. 2 Transcriptional activation by homoeoproteins via specific target sequences. *a*, In the absence of an SV40 enhancer. One of eight different OVEC plasmids bearing various targets (Fig. 1*b*) were co-transfected in pairs with one of five different expression plasmids bearing potential *trans*-activators (a, pCA expressing *Abd-B*⁺; u, pCU expressing *Ubx*; o, pCO expressing OTF-2; c1, control plasmid carrying a non-functional glucocorticoid receptor fragment¹⁵; c2, pCD containing a *Drosophila Deformed* cDNA³⁰, another homoeotic gene product: this plasmid served as a second control as we could not detect a trace of *Deformed* protein in transfected HeLa cells by antibody staining. Correctly initiated β -globin transcripts (5' ends) are indicated by arrowheads: signals from the target-bearing OVEC (s) and from the reference plasmid OVEC-REF (r; in each case, same gel as s, but shorter exposure). All homoeoproteins (a, u, o) activate the B and E targets as single copies and, more strongly, as double copies (2B, 2E) as well as the Octa1 target, but not OVEC, 4GRE or OVEC Bsub. *b*, In the presence of an SV40 enhancer. In this series, all OVEC plasmids contain an SV40 enhancer (see Fig. 1) which increases the reporter gene signal (hence high s:r ratios). As in (*a*), the homoeoproteins activate OVEC B, OVEC E and, to a lesser extent, OVEC Octa2 (u activation hardly detectable), but not an OVEC plasmid without insertion of a functional target (ov). The SV40 enhancer reveals endogenous activators which act on the B and E targets (compare c1 lanes of ov to c1 lanes of B and E).

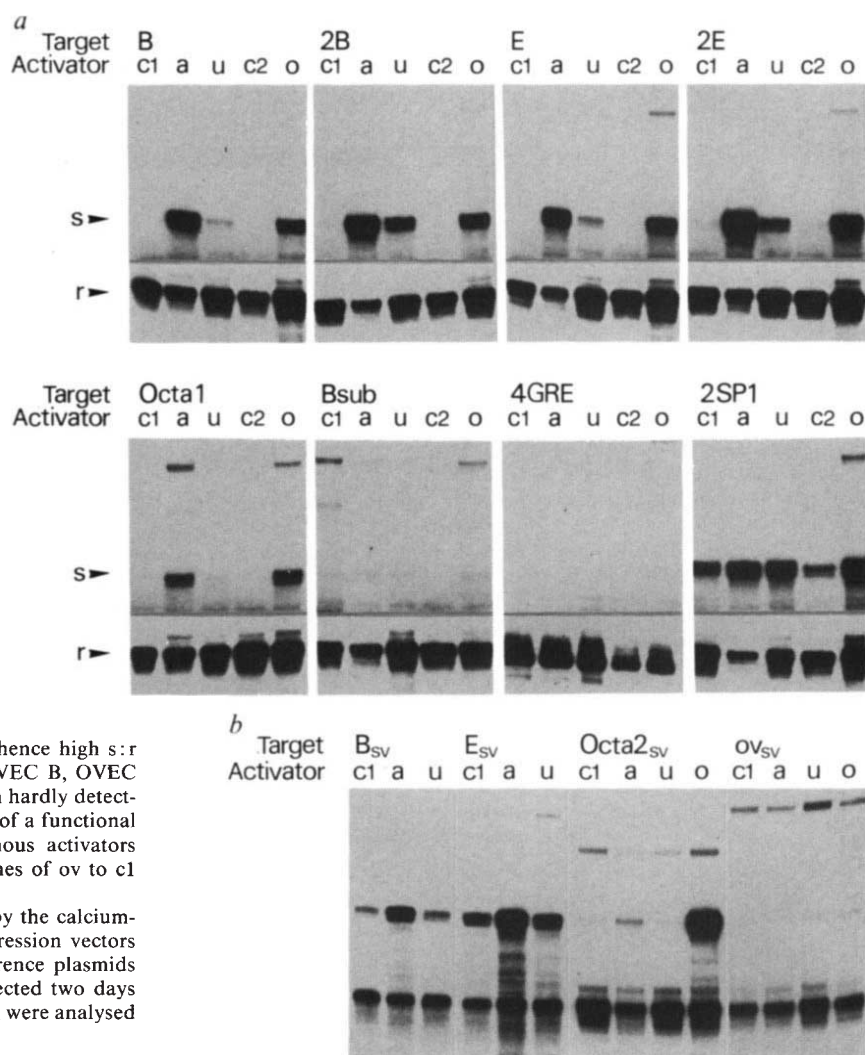
Methods. Transfections of HeLa cells were done by the calcium-phosphate method as previously described¹¹ (expression vectors 5 μ g, OVEC reporter plasmids 10 μ g, OVEC reference plasmids 0.5 μ g per transfected petri dish). Cells were collected two days later, total RNA was prepared and samples of 20 μ g were analysed by the SP6-RNase protection method as before¹⁹.

scription factors²⁰ such as SP1. The *Ubx* protein (lanes marked u) showed essentially the same activity as *Abd-B*⁺, albeit to a lesser extent (up to 23-fold). The activity of the *Drosophila* homoeoproteins is similar to the transcriptional induction obtained by OTF-2 (lane marked o) upon co-transfection with OVEC Octa1 (13-fold). We conclude that the proteins encoded by *Drosophila* homoeotic genes are transcriptional activators and that they can function in the absence of any other *Drosophila* protein in mammalian cells. It is almost certain that this activity is mediated by direct DNA binding, as *Abd-B*⁺ and *Ubx* protein generated by *in vitro* translation or by expression in *Escherichia coli* specifically bind to the B and E sequences (not shown).

Interestingly, the *Abd-B*⁺ and *Ubx* proteins (lanes marked a, u) are also capable of activating transcription via the octamer binding site Octa1 (up to 14-fold). Conversely, OTF-2 (lanes marked o) acts on the *Drosophila* targets B and E (up to 23-fold). Thus, all three homoeoproteins are functionally equivalent in that they stimulate transcription by acting on the same set of target sequences.

We asked whether these sequences were related in any way. Systematic computer searches²¹ established that the octamer motif contained in OVEC Octa1 convincingly matches a sequence within OVEC B as well as in OVEC E (Fig. 3). The latter match is shifted by three residues with respect to the homoeoprotein binding consensus sequence^{9,10}; however, naturally occurring copies of this sequence in *Drosophila* promoters are typically found in tandem repeats^{9,10} and experimental tests mostly involved such repeats⁹.

The low levels of B and E OVEC expression observed in the absence of co-transfected *trans*-activators (lanes c1, c2) indicate



that the level and/or activity of putative endogenous homoeoproteins is much lower compared to that of the endogenous SP1 transcription factor. If we assay these targets in OVEC plasmids bearing an SV40 enhancer, the level of reporter gene expression is much enhanced in the absence of added *trans*-activators, but stimulated even further in their presence (Fig. 2*b*). This enhancement is dependent on the presence of a functional target, suggesting that endogenous HeLa and co-transfected homoeoproteins can mediate²⁰, if bound adjacent to the TATA-box, long-distance stimulation from the SV40 enhancer. One of the endogenous HeLa homoeoproteins might

Octa1-C	GAA	ATGCAAATTA	CCC
Octa2-C	ATA	ATGCAAATTA	GAG
B	ATT	ATGCAAATTA	TCA
E	TAA	ATGCAAATTA	AAT
conserved	--A	ATGCAAATTA	---
Octa Consensus		ATGCAAATnA	

Fig. 3 Homologies between the homoeoprotein targets. A computer search²¹ identified a sequence stretch in B as well as in E with significant similarity to the decamer sequences contained in Octa1 and Octa2 (identical residues are highlighted; complements of Octa1 and Octa2 are given). The consensus sequence derived from a compilation of octamer binding sites (E. Schreiber, personal communication) is given below. This sequence is out of register by three residues with respect to the homoeoprotein binding consensus sequence^{9,10} (see text). The only other similarities found in this search were between Bsub and 4GRE in the linker region and between Bsub and 2SP1.

be the ubiquitous octamer transcription factor OTF-1 that is presumably related to OTF-2 as it binds to similar sequences²².

Our results suggest that the Abd-B^r protein and OTF-2 bind to the *Drosophila* and the octamer target sequences with different affinities: Abd-B^r activates OVEC B and E better than OTF-2, whereas OTF-2 activates OVEC Octal better than Abd-B^r (Fig. 2a, compare lanes a and o in B, E and Octal). In contrast, Ubx activity always parallels Abd-B^r activity, thus there is no evidence for differential binding affinities between these two proteins. Differential DNA-binding of distantly related homoeoproteins⁹ is assumed to reflect structural differences in the part of the homoeodomain that corresponds to the DNA recognition helix²³. Indeed, the Abd-B^r and the Ubx protein are essentially identical^{14,23}, whereas the OTF-2 protein differs considerably from the others¹¹ in this region.

Our results suggest that the different homoeoproteins have different activation potentials, Abd-B^r being a more potent activator than Ubx. We attempted to compare the levels of Abd-B^r and Ubx expression in the transfected HeLa cells by antibody staining of whole cells and of cell extract western blots and found that Ubx protein is expressed at somewhat higher levels than Abd-B^r (data not shown). Thus, it seems that the Abd-B^r protein might be inherently more active than the Ubx protein as regards *trans*-activation.

Genetic analysis suggested that homoeotic gene products can act as activators or as repressors. For example, in those cells where Ubx autocatalyses its own expression³, it acts as a repressor on *Antennapedia* expression (G. Tremml and M. B., manuscript in preparation). It was found that Abd-B^r functions to repress Abd-B^m (ref. 24) and other homoeotic genes such as Ubx (ref. 25), but Abd-B^r may also have an activating ('morphogenetic') function²⁴. Although it is possible that some of these interactions are not direct, there are precedents for eukaryotic regulators that can directly activate or repress transcription, depending on the target promoter^{26,27}. We imagine that an accessory protein in *Drosophila*, a DNA-binding protein in this case, may convert an activating homoeoprotein into a repressing one by masking its activation domain, similar to the way in which yeast GAL80 masks the yeast GAL4 activator²⁸.

We thank Hugh Pelham for discussion and comments on the manuscript, Sandro Rusconi for plasmids and helpful advice and Michael O'Connor and Welcome Bender for the Ubx cDNA clone. This work was supported by the Swiss NSF.

The POU domain is a bipartite DNA-binding structure

Richard A. Sturm & Winship Herr

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

The POU domain¹ (pronounced 'pow') is a highly charged 155–162-amino-acid (aa) region of sequence similarity contained within three mammalian transcription factors. Pit-1 (ref. 2), Oct-1 (ref. 3) and Oct-2 (ref. 4), and the product of the nematode gene *unc-86* (ref. 5) which is involved in determining neural cell lineage. This domain consists of two subdomains, a C-terminal homoeo domain and an N-terminal POU-specific region separated by a short nonconserved linker; the sequence relationship shows that the POU homoeo domains form a distinct POU-related family. In the ubiquitous and lymphoid-specific octamer-motif binding proteins Oct-1 and Oct-2, the POU domain is sufficient for sequence-specific DNA binding^{3,4}. Homoeobox domains contain a helix-turn-helix DNA-binding motif^{6,7}, first identified in bacterial repressors⁸. The helix-turn-helix region of the POU domain is important for DNA binding^{3,9} and, in other classes of homoeo-containing proteins, the entire homoeo domain is sufficient for DNA binding^{10–12}; thus the new POU-specific region could be involved in other functions such as protein-protein interactions. Nevertheless, we show here that in fact the POU domain is a novel bipartite DNA-binding structure in which the POU homoeo and POU-specific regions form two subdomains that are both required for DNA binding but are held together by a flexible linker.

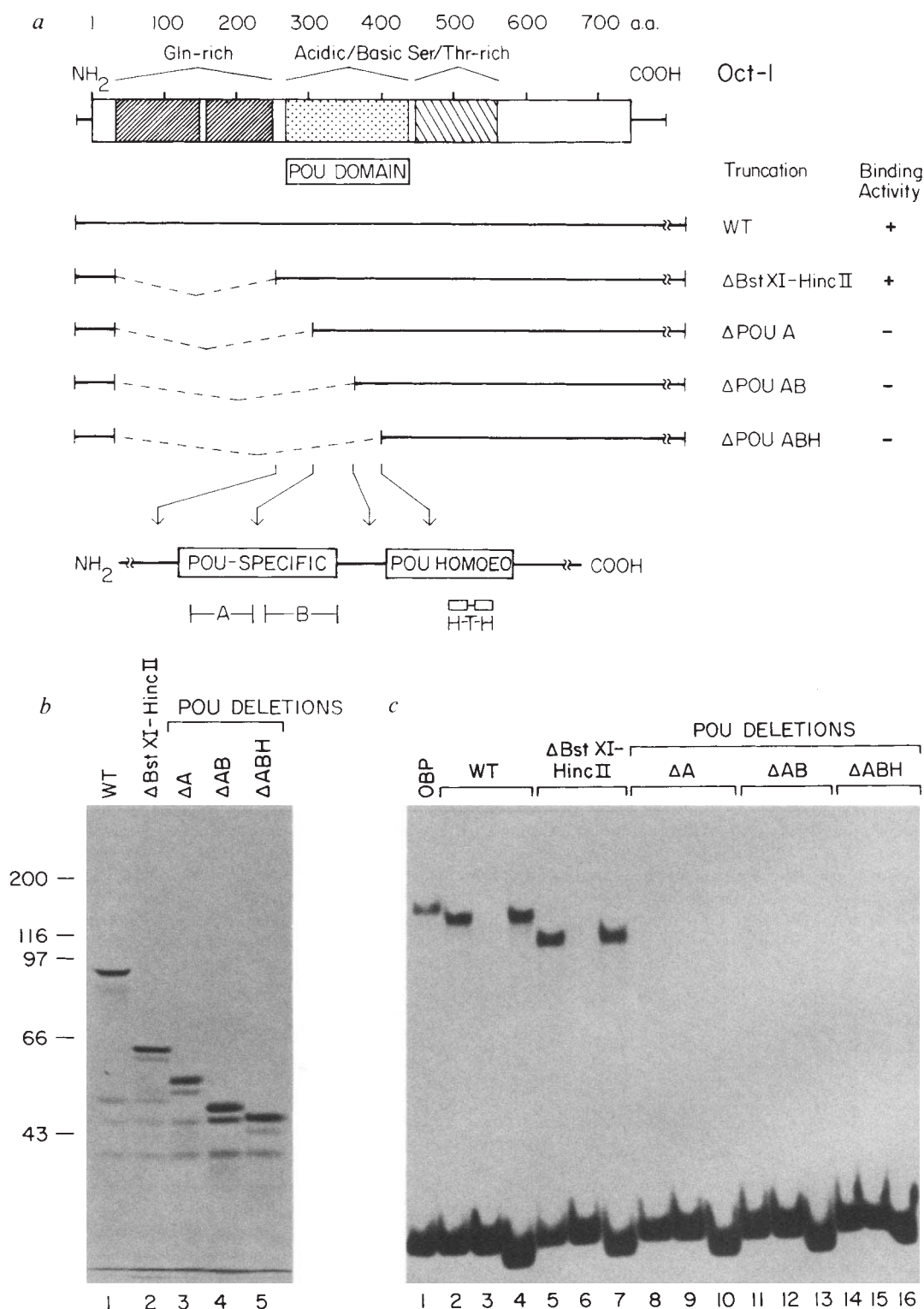
We used a cDNA copy of the *oct-1* gene³ to study the Oct-1 DNA binding domain. Oct-1 is probably the same as the ubiquitous octamer-binding transcription and adenovirus replication factor variously called NF-A1 (ref. 13), OTF-1 (ref. 14), NFIII (ref. 15), and OBP100 (ref. 16) (see ref. 17). Figure 1 shows the effect of progressively larger N-terminal deletions of the Oct-1 POU domain on sequence-specific DNA binding in a gel retardation assay. The top of the figure shows the position of the POU domain in the Oct-1 coding region, flanked by glutamine and serine/threonine-rich regions³. The position of the POU-specific and POU homoeo subdomains is shown below the deletion map. The comparison of the four POU proteins showed that the POU-specific region or 'box' may be further subdivided by the degree of sequence similarity into A and B subregions¹ (Fig. 1a). Therefore, to define the N-terminal boundary of the Oct-1 DNA-binding domain, we extended the N-terminal Δ BstXI-HincII deletion³ to include progressively the POU A box (Δ POU A), the entire POU-specific box (Δ POU AB), and additionally the N-terminal region of the POU homoeo domain (Δ POU ABH). In each of these mutants the initiation codon and 5' untranslated region is kept constant by fusing 23 aa from the predicted N-terminus of the Oct-1 protein³, via a 2-aa leu-glu linker, to the different deletion end points. The proteins resulting from *in vitro* transcription and translation of the wild-type and deleted *oct-1* constructs are shown in Fig. 1b.

The DNA-binding activity of each Oct-1 protein was tested in a gel retardation assay with three different DNA fragments containing either the wild-type SV40 site-I octamer motif (Octal), a defective SV40 site-I motif containing two point mutations (*dpm8*), or a herpes simplex virus TAATGARAT (R = purine) motif. The Oct-1 protein binds to very degenerate octamer motifs^{3,16,18}. The SV40 site-I octamer motif and the TAATGARAT motif used here are the two most dissimilar Oct-1 binding sites identified: they match at only four out of 14 positions¹⁸. Figure 1c shows that the two proteins containing the entire POU domain (WT and Δ BstXI-HincII) can bind to the different SV40 and TAATGARAT sites but not to the mutant SV40 octamer site (lanes 2–7). Deletion of the POU A box,

Received 14 October; accepted 31 October 1988.

1. Akam, M. *Development* **101**, 1–22 (1987).
2. Garcia-Bellido, A. in *Cell Patterning, CIBA Foundation Symposium* Vol. 29 (ed. Brenner, S.) 161–182 (New York Associated Scientific, 1975).
3. Bienz, M. & Tremml, G. *Nature* **333**, 576–578 (1988).
4. Hafen, E., Levine, M. & Gehring, W. *Nature* **307**, 287–289 (1984).
5. Struhl, G. & White, R. A. H. *Cell* **43**, 507–519 (1985).
6. Shepherd, J. C. W., McGinnis, W., Carrasco, A. E., DeRobertis, E. M. & Gehring, W. *Nature* **310**, 70–71 (1984).
7. Laughon, A. & Scott, M. P. *Nature* **310**, 25–31 (1984).
8. Desplan, C., Theis, J. & O'Farrell, P. H. *Nature* **318**, 630–635 (1985).
9. Desplan, C., Theis, J. & O'Farrell, P. H. *Cell* **55**, 1081–1090 (1988).
10. Hoey, T. & Levine, M. *Nature* **332**, 858–861 (1988).
11. Müller, M. M., Ruppert, S., Schaffner, W. & Matthias, P. *Nature* **336**, 544–551 (1988).
12. Ptashne, M. *Nature* **335**, 683–689 (1988).
13. O'Connor, M. B., Binari, R., Perkins, L. A. & Bender, W. *EMBO J.* **7**, 435–445 (1988).
14. DeLorenzi, M. *et al.* *EMBO J.* **7**, 3223–3231 (1988).
15. Severne, Y., Wieland, S., Schaffner, W. & Rusconi, S. *EMBO J.* **7**, 2503–2508 (1988).
16. Biggin, M. D. & Tjian, R. *Cell* **53**, 699–711 (1988).
17. Bienz, M. *et al.* *Cell* **53**, 5567–5576 (1988).
18. Westin, G., Gerster, T., Müller, M. M., Schaffner, G. & Schaffner, W. *Nucleic Acids Res.* **15**, 6787–6798 (1987).
19. Höller, M., Westin, G., Jiricny, J. & Schaffner, W. *Genes Dev.* **2**, 1127–1135 (1988).
20. Green, M. R., Treisman, R. & Maniatis, T. *Cell* **35**, 137–148 (1983).
21. Staden, R. *Nucleic Acids Res.* **10**, 2951–2961 (1982).
22. Fletcher, C., Heintz, N. & Roeder, R. G. *Cell* **51**, 773–781 (1987).
23. Otting, G. *et al.* *EMBO J.* (in the press).
24. Casanova, J., Sanchez-Herrero, E. & Morata, G. *Cell* **47**, 627–636 (1986).
25. Casanova, J. & White, R. A. H. *Development* **101**, 117–122 (1987).
26. Shore, D. & Nasmyth, K. *Cell* **51**, 721–732 (1987).
27. Sakai, D. D. *et al.* *Genes Dev.* **2**, 1144–1154 (1988).
28. Ma, J. & Ptashne, M. *Cell* **50**, 137–142 (1987).
29. Weinzierl, R., Axton, J. M., Ghysen, A. & Akam, M. *Genes Dev.* **1**, 386–397 (1987).
30. Jack, T., Regulski, M. & McGinnis, W. *Genes Dev.* **2**, 635–651 (1988).
31. Saari, G. & Bienz, M. *EMBO J.* **6**, 1775–1779 (1987).

Fig. 1 N-terminal deletion mapping of the Oct-1 DNA-binding domain. **a**, The structure of the predicted 743-aa Oct-1 *in vitro* translation product³ (boxed) is illustrated with the shaded portions indicating regions of the Oct-1 protein that are rich in either glutamine, acidic and basic, or serine and threonine residues. The centre of the panel shows the Oct-1-coding sequences removed in each deletion with the names and DNA-binding activities indicated to the side. Each deletion begins with residue 24 (inclusive) and the last residue deleted is 269 (Δ BstXI-HincII), 316 (Δ POU A), 369 (Δ POU AB), or 400 (Δ POU ABH). Each mutant has a leu-glu linker replacing the deleted sequences. The position of the POU domain is indicated at the top and an enlargement of the POU domain is shown at the bottom of the panel to indicate the position of the deletions relative to the POU-specific A and B, and the POU homoeo subdomains. The position of the helix-turn-helix (H-T-H) motif is indicated. **b**, The [³⁵S]methionine-labelled *in vitro* translation products generated from the wild-type (WT) (lane 1), Δ BstXI-HincII (lane 2), Δ POU A (lane 3), Δ POU AB (lane 4), and Δ POU ABH (lane 5) *oct-1* templates were resolved on a 10% SDS-polyacrylamide gel and visualized by autoradiography. **c**, The *in vitro* Oct-1 translation products shown in panel **b** were tested for sequence-specific DNA-binding activity in a gel retardation assay using the following ³²P-labelled binding sites: the wild-type SV40 site I (Oct-1) sequence AGT-ATGCAAAGCAT (lanes 1, 2, 5, 8, 11 and 14); the *dpm8* double mutant Oct-1 site AGTAaGgAAAGCAT (lanes 3, 6, 9, 12 and 15); and the herpes simplex virus TAATGARAT motif GGCATCTCATACC (lanes 4, 7, 10, 13 and 16). (The octamer-related sequences are underlined. Note that the TAATGARAT motif is on the opposite strand; see ref. 3, and ref. 18 for a complete description of the probes.) Lane 1 shows authentic HeLa cell OBP100 (Oct-1)-binding activity with the SV40 site-I probe.



Methods. The internal *oct-1* deletions were generated by oligonucleotide-directed mutagenesis as described previously¹⁶ using single-stranded pBS_{oct-1}⁺ (ref. 3) DNA produced from the *ung*⁻*dut*⁻ *E. coli* strain BW313 and the following oligomers: Δ POU A, ACAGGCACACAAACCCTCGAGTATGGAAATGACTTC; Δ POU AB, ACAGGCACACAAACCCTCGAGTCTCCAGGAATTGAG; Δ POU ABH, ACAGGCACACAAACCCTCGAGAATCAAAAGCCTACC. *In vitro* transcriptions and translations, and gel retardation analysis were performed as described previously^{3,16}. One μ l of *in vitro* synthesized Oct-1 protein was pre-incubated with 1 μ g of poly (dI-dC). poly(dI-dC) in a 10 μ l reaction for 10 min before addition of ³²P radio-labelled probe as described³.

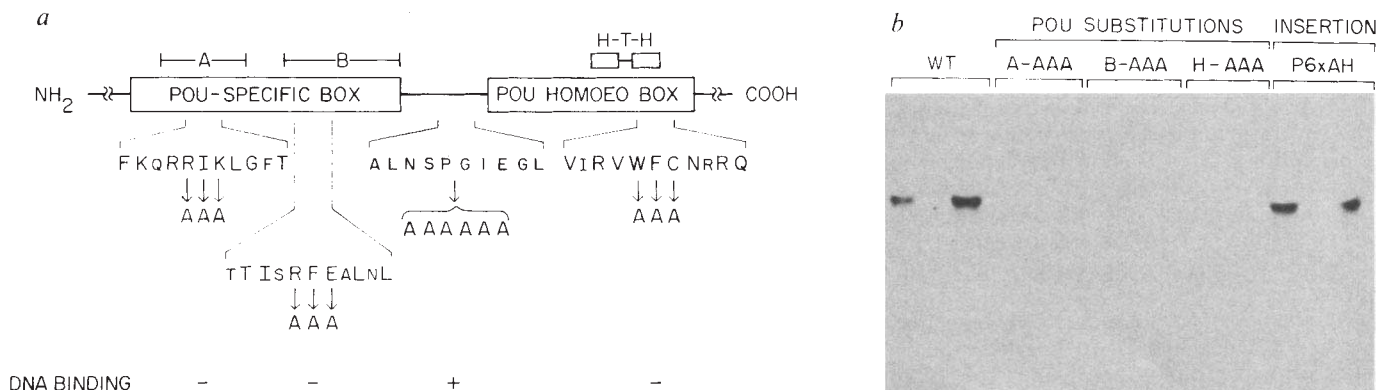


Fig. 2 The Oct-1 POU domain requires functional POU-specific and POU homoeo subdomains but the linker is flexible. *a*, The Oct-1 POU domain is shown with the position of the POU-specific and POU homoeo subdomains indicated. The position of the POU-specific A and B boxes¹ and the homoeo helix-turn-helix (H-T-H) motif is shown above the boxed regions. Short stretches of the Oct-1 amino-acid sequence at the sites of mutagenesis are shown in one-letter code, with those residues that are the same among all four POU proteins indicated by the larger letters. The three positions mutated to alanine residues in each mutant are shown and the position at which six alanines were inserted is indicated by the bracket. The effect of the mutations on DNA-binding activity is indicated at the bottom of the panel. *b*, Full length *in vitro* translated mutant Oct-1 proteins shown in *a* were tested for DNA-binding activity in a gel retardation assay as described in Fig. 1. The proteins used were wild-type *in vitro* translated protein (lanes 1-3), the A-box mutant A-AAA (lanes 4-6), the B-box mutant B-AAA (lanes 7-9), the homoeobox helix-turn-helix mutant H-AAA (lanes 10-12), and the six alanine insertion mutant in the POU linker P6xAH (lanes 13-15). The probes alternate in triplets (left to right) wild-type SV40 site I, the *dpm8* mutant SV40 site I, and the TAATGARAT motif (see Fig. 1). The methods are as described in Fig. 1. The levels of *in vitro* translated wild-type and mutant Oct-1 proteins were shown to be equivalent by SDS-polyacrylamide gel electrophoresis before using 2 μ l of *in vitro* translation extract for the gel retardation analysis. The oligonucleotides used for mutagenesis were as follows:

A-AAA, CAAACAAAGAGCAGCCGACITGGATTC;
 B-AAA, TACCATCTCTGCAGCTGCAGCCTTGAAC;
 H-AAA, GATTCGTGTTGCGGCCGCTAACCGCCGC;
 P6xAH, CTGAATTCTCCAGCTGCCGACGCGCTGCCGAATTGAGGGC.

however, abolishes detectable binding in this assay to either binding site (lanes 8-10). The entire POU-specific box deletion Δ POU AB and the Δ POU ABH deletion also do not bind detectably to the Oct-1 binding sites (lanes 11-16). Therefore, the POU homoeo domain is insufficient for effective DNA binding.

The effect on DNA binding of more subtle mutations in the POU domain was tested by creating triple alanine substitutions at positions that are identical in all four of the POU proteins. Three regions were mutated as shown in Fig. 2a: the POU-specific A box (A-AAA) and B box (B-AAA), and the putative DNA-recognition helix of the POU homoeo domain (H-AAA). We chose alanine substitutions because this residue is less likely to affect the backbone structure of the protein. Assay of each of these three mutant Oct-1 proteins as full length *in vitro* translation products by gel retardation (Fig. 2b, lanes 4-12) shows that each protein is defective for binding to both of the Oct-1 binding sites. Thus both the POU-specific and the POU homoeo domains are required for recognition of the two very different DNA sequences. But the linker that holds the two subdomains together is flexible because insertion of six alanine residues within this region (Fig. 2a, P6xAH) does not have a detectable effect on DNA binding (lanes 13-15). These results lead to a picture of the POU domain as a bipartite DNA-sequence-recognition domain with POU-specific and POU homoeo domains as substructures.

These studies do not address how the POU-specific subdomain is involved in DNA-sequence recognition. The POU homoeo subdomain contains a helix-turn-helix motif and therefore probably contacts the DNA directly. The structure of the POU-specific domain is not readily predictable; although it has the potential to form multiple α helices, it does not contain a recognizable helix-turn-helix motif. The POU-specific domain could serve to stabilize the homoeo domain and not directly

contact the DNA. Nevertheless, we believe both POU subdomains make direct DNA contacts because Oct-1 probably binds as a monomer (unpublished data; see ref. 4 for Oct-2 monomer binding) and makes many major and minor groove contacts with the DNA¹⁹ over 11-14 base pairs¹⁸. Such an extensive number of contacts is not expected for a single helix-turn-helix binding domain²⁰. Using an assay different from gel retardation (for example, DNaseI footprinting) and an excess of each subdomain it may be possible to determine whether each half of the POU domain can bind directly to DNA.

An interesting feature of the POU domain is the flexible linker suggesting that the two POU subdomains do not need to be precisely held together. It is apparently important, however, that the two subdomains be attached to one another because cotranslation of the two individual subdomains did not produce detectable gel retardation activity (data not shown). The flexible hinge may allow flexibility in the protein/DNA interaction such that the DNA-binding site itself could expand or contract. For example, the very different octamer (ATGCAAAT) and TAATGARAT motifs can be made identical if a C:G base pair is inserted into the TAATGARAT motif (TAATGCAAAT).

Although the studies described here show that the POU domain, which is found in a nematode cell-lineage gene and both ubiquitous and cell-specific mammalian transcription factors, is involved in DNA binding, this domain may also be involved in other functions such as interactions with other transcription factors. A paradigm for such interactions is the DNA-binding domain of λ repressor which also makes direct contact with the RNA polymerase²¹.

The POU domain represents a new class of DNA-binding structure. Different classes of DNA-binding domains may confer unique properties on the transactivator. For example, the multiple Zn²⁺-finger DNA-binding structures of TFIIIA, which binds to the intragenic 5S RNA promoter, may be directly

responsible for its ability to remain bound to the DNA as the RNA polymerase transcribes through the gene^{22,23}. In the case of the POU domain, the bipartite structure may envelope the DNA such that the protein surrounds the DNA. Such a structure could avoid stereo-specific constraints on interactions with other transcription factors or allow interactions with multiple factors simultaneously on opposite sides of the helix. Such properties may aid the formation of stable transcription complexes as may be expected in generally expressed genes (small nuclear RNAs or histones for Oct-1) or genes that confer a stable differentiated phenotype.

We thank M. Cleary for DNA sequence analysis of the mutants, J. Duffy and D. Greene for expeditious artwork, M. Iacovacci for rapid production of oligonucleotides, T. Grodzicker, H. R. Horvitz, and J. Pflugrath for valuable discussions, B. Ondek for reading the manuscript, and U. Grossniklaus and W. Gehring's laboratory for hospitality during preparation of the manuscript. R.A.S. is a recipient of a Cancer Research Institute Fellowship, New York, and W.H. is a Rita Allen Foundation Scholar. This work was supported by a grant from the National Cancer Institute.

Received 25 October; accepted 2 November 1988.

- Herr, W. *et al. Genes Dev.* **2**, 1513-1516 (1988).
- Ingraham, H. A. *et al. Cell* **55**, 519-529 (1988).
- Sturm, R. A., Das, G. & Herr, W. *Genes Dev.* **2**, 1582-1599 (1988).
- Clerc, R. G., Corcoran, L. M., LeBowitz, H. H., Baltimore, D. & Sharp, P. A. *Genes Dev.* **2**, 1570-1581 (1988).
- Finney, M., Ruvkun, G. & Horvitz, H. R. *Cell* (in the press).
- Laughon, A. & Scott, M. P. *Nature* **310**, 25-31 (1984).
- Otting, G. *et al. EMBO J.* (in the press).
- Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M. & Pabo, C. O. *Nature* **298**, 447-451 (1982).
- Ko, H.-S., Fast, P., McBride, W. & Staudt, L. M. *Cell* **55**, 135-144 (1988).
- Desplan, C., Theis, J. & O'Farrell, P. H. *Cell* **54**, 1081-1090 (1988).
- Hoey, T., Warrior, R., Manak, J. & Levine, M. *Molec. cell. Biol.* **8**, 4598-4607 (1988).
- Muller, M. *et al. EMBO J.* (in the press).
- Staudt, L. M. *et al. Nature* **323**, 640-643 (1986).
- Fletcher, C., Heintz, N. & Roeder, R. G. *Cell* **51**, 773-781 (1987).
- Prujin, G. J. M., van Driel, W. & van der Vliet, P. C. *Nature* **322**, 656-659 (1986).
- Sturm, R., Baumruker, T., Franza, B. R. Jr & Herr, W. *Genes Dev.* **1**, 1147-1160 (1987).
- O'Neill, E. A. *et al. Science* **241**, 1210-1213 (1988).
- Baumruker, T., Sturm, R. & Herr, W. *Genes Dev.* **2**, 1400-1413 (1988).
- Prujin, G. J. M., van Miltenburg, R. T., Claessens, J. A. J. & van der Vliet, P. C. *J. Virol.* **62**, 3092-3102 (1988).
- Anderson, J. E., Ptashne, M. & Harrison, S. C. *Nature* **326**, 846-852 (1987).
- Hochschild, A., Irwin, N. & Ptashne, M. *Cell* **32**, 319-325 (1983).
- Miller, J., McLachlan, A. D. & Klug, A. *EMBO J.* **4**, 1609-1614 (1985).
- Wolffe, A. P., Jordan, E. & Brown, D. D. *Cell* **44**, 381-389 (1986).

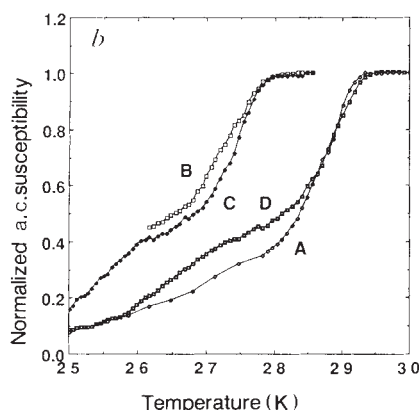
Erratum

The oxygen isotope effect in $\text{Ba}_{0.625}\text{K}_{0.375}\text{BiO}_3$

D. G. Hinks, D. R. Richards, B. Dabrowski, D. T. Marx & A. W. Mitchell

Nature **335**, 419-421 (1988).

THE correct version of Fig. 1b is shown below.



Copies of articles from this publication are now available from the UMI Article Clearinghouse.

For more information about the Clearinghouse, please fill out and mail back the coupon below.

UMI Article Clearinghouse

Yes! I would like to know more about UMI Article Clearinghouse. I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem
☐ Other (please specify) _____
☐ I am interested in sending my order by mail.
☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____

Title _____

Institution/Company _____

Department _____

Address _____

City _____ State _____ Zip _____

Phone (_____) _____

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106